

REMARKS

Claims 53-58, 71-73 and 86-94 are pending. Claims 53-58, 88 and 90-94 are amended, and new claim 95 is added. No new matter has been added by the amendments.

Election/Restrictions

The Applicants note the finality of the Restriction Requirement. The Applicants however respectfully request rejoinder of at least the species SEQ ID NOs: 22 and 23 in view of the submissions herein.

Specification

The disclosure is objected to for containing an embedded hyperlink and/or other form of browser-executable code at p. 54-55. In response, the paragraph spanning p. 54-55 has been amended to remove the browser-executable code.

Rejections Under 35 U.S.C. § 112

I. Claims 53-58, 71-73 and 86-94 are rejected under 35 U.S.C. 112, first paragraph, as allegedly failing to comply with the written description requirement. More specifically, while the Examiner acknowledges that the specification describes an actual reduction to practice of SEQ ID NO: 24, the Examiner asserts that the claims are drawn to a large genus of variants of SEQ ID NO: 24, and that the Applicants were allegedly not in possession of this genus with respect to fragments or variants or substantially identical sequences.

Claims 53-55, which are the only pending independent claims, have been amended without prejudice or disclaimer and without acquiescence to the Examiner's assertions, to recite that the claimed polypeptide comprises an amino acid sequence having at least 75% sequence identity to SEQ ID NO: 24 and to recite an "immunogenic fragment." Support for these amendments may

be found throughout the specification at for example page 13, lines 6-28; page 18, line 26 to page 19, line 6; or page 30, line 12, to page 31, line 23. The Applicants respectfully submit that the specification describes at least three variants of SEQ ID NO: 24, *i.e.*, NleA polypeptides from enteropathogenic *E. coli* (EPEC) and from *C. rodentium*, as well as from enterohemorrhagic *E. coli* (EHEC), which fall within the presently claimed sequence identity, and therefore describes a representative number of species with respect to the claimed genus. Furthermore, recitation of an “immunogenic fragment” clarifies that only fragments capable of eliciting an immune response are contemplated. Accordingly, this rejection should be withdrawn.

II. Claims 53-58, 71-73 and 86-94 are rejected under 35 U.S.C. 112, first paragraph, as allegedly failing to comply with the enablement requirement. More specifically, while the Examiner acknowledges that the specification is enabling for a method for eliciting an immune response against an enterohemorrhagic *E. coli* (EHEC) or SEQ ID NO: 24 (which is characterized as a component of enterohemorrhagic *E. coli* O157:H7), or for reducing colonization or shedding of EHEC, or for treating EHEC infection, in an animal by administering an effective amount of a composition or cell culture supernatant including a polypeptide which comprises the amino acid sequence set forth in SEQ ID NO: 24, the Examiner alleges that other aspects of the claimed invention are not enabled. In particular, the Examiner alleges that the specification does not reasonably provide enablement for preventing infection by EHEC in an animal by administering an effective amount of a composition or cell culture supernatant including a polypeptide which comprises the amino acid sequence set forth in SEQ ID NO: 24. The Examiner also alleges that the specification does not reasonably provide enablement for the claimed methods with respect to any other A/E pathogen, or component thereof, or for another polypeptide comprising an amino acid sequence substantially identical to the sequence of SEQ ID NO: 24 or a fragment or variant thereof.

Claims 53-55, which are the only pending independent claims, have been amended without prejudice or disclaimer and without acquiescence to the Examiner's assertions, to recite that the claimed polypeptide comprises an amino acid sequence having at least 75% sequence identity to

SEQ ID NO: 24 and to recite an “immunogenic fragment.” In addition, the present claims do not recite prevention of infection.

As indicated herein, the specification describes at least three variants of SEQ ID NO: 24, *i.e.*, NleA polypeptides from EPEC and from *C. rodentium* and therefore describes a representative number of species with respect to the claimed genus. Furthermore, one of ordinary skill in the art would also be able to readily identify NleA variants as, for example, evidenced in the enclosed publication by Creuzburg and Schmidt (J. Clin. Microbiol. 2498-2507, 2007) in which a large number of NleA variants were detected after the initial identification of NleA as a virulence factor by the inventors of the above-referenced application.

The Applicants also respectfully submit that the specification demonstrates the effect of NleA in a *C. rodentium* mouse model of disease and that one of ordinary skill in the art would be able to readily apply the claimed methods to A/E pathogens, as claimed. Accordingly, one of ordinary skill in the art would be able to readily identify variants and immunogenic fragments of NleA proteins, and to use the claimed methods in connection with A/E pathogens, and the Applicants respectfully request withdrawal of this rejection.

Rejections Under 35 U.S.C. § 102

I. Claims 53-58, 71-72 and 86-94 are rejected under 35 U.S.C. 102(b) as allegedly anticipated by Finlay *et al.* (WO 02/053181) as evidenced by Hideo *et al.* (JP20023550742A2, partial translation and sequence alignment attached as Appendix B and Appendix A, respectively, of the Office Action).

More specifically, the Examiner alleges that Finlay *et al.* teach methods for eliciting an immune response against an A/E pathogen or component thereof, or for reducing colonization of an A/E pathogen, or of reducing shedding (thus allegedly treating an infection by an A/E pathogen) in an animal by administering an effective amount of a composition comprising a culture supernatant. The Examiner further alleges that Hideo *et al.* teach that *E. coli* EHEC

O157:H7 makes a protein comprising the sequence of SEQ ID NO: 24. The Office Action further alleges that the culture supernatant of Finlay *et al.* is prepared from *E. coli* EHEC O157:H7 under identical conditions as SEQ ID NO: 24 of the instant specification.

The Examiner therefore concludes that the culture supernatant of Finlay *et al.* is a composition or culture supernatant which comprises a polypeptide which comprises an amino acid sequence substantially identical to the sequence of SEQ ID NO: 24 and inherently comprises 20% of the cell protein present in the composition. The Examiner is respectfully requested to clarify “inherently comprises 20% of the cell protein present in the composition” in the context of this rejection.

To support a rejection under § 102, a single prior art reference must describe each and every element, either expressly or inherently, of the rejected claims (MPEP § 2131). In the present case, claims 53-55, which are the only pending independent claims, have been amended without prejudice or disclaimer and without acquiescence to the Examiner’s assertions, to recite that the claimed polypeptide is “isolated.” The term “isolated” as defined in the specification refers to a compound that is “separated from the components that naturally accompany it” (see, for example, the specification at page 10, lines 20-21). The Applicants respectfully submit that Finlay *et al.* do not teach methods relating to an “isolated” polypeptide comprising an amino acid sequence having at least 75% sequence identity to the sequence of SEQ ID NO: 24 as claimed and therefore do not anticipate the claimed invention.

II. Claims 53-58, 71-72, 86, and 88-94 are rejected under 35 U.S.C. 102(b) as allegedly anticipated by Wright (US 5,730,989, 3/24/98) as evidenced by Hideo *et al.* (*supra*).

More specifically, the Examiner alleges that Wright disclose a method for eliciting an immune response against *E. coli* O157:H7 or component thereof, in an animal by administering to the animal an effective amount of inactivated *E. coli* O157:H7. The Examiner further alleges that the *E. coli* O157:H7 of Wright is a composition that comprises a polypeptide which comprises an amino acid sequence substantially identical to the sequence of SEQ ID NOs: 24, as evidenced by

Hideo *et al.*, and that Wright disclose a method for treating *E. coli* infection, thus treatment of the *E. coli* infection will allegedly result in reduction in colonization and shedding of *E. coli* in an animal.

As indicated herein, to support a rejection under § 102, a single prior art reference must describe each and every element, either expressly or inherently, of the rejected claims (MPEP § 2131). In the present case, claims 53-55, which are the only pending independent claims, have been amended without prejudice or disclaimer and without acquiescence to the Examiner's assertions, to recite that the claimed polypeptide is "isolated." The Applicants respectfully submit that Wright *et al.* do not teach methods relating to use of an "isolated" polypeptide comprising an amino acid sequence having at least 75% sequence identity to the sequence of SEQ ID NO: 24 as claimed and therefore do not anticipate the claimed invention.

III. Claims 53-55, 71-72, 86 and 90 are rejected under 35 U.S.C. 102(b) as allegedly anticipated by Hideo *et al.* (*supra*) as evidenced by Wright *et al.* (*supra*).

More specifically, the Examiner alleges that Hideo *et al.* disclose a method of eliciting an immune response against *E. coli* O157:H7 by administering an effective amount of a composition for inducing an immune response against *E. coli* O157:H7 comprising a protein 100% identical to SEQ ID NO: 24. The Examiner further alleges that Hideo *et al.* also disclose treating an infection by *E. coli* O157:H7 using the composition and concludes that treatment of the *E. coli* infection will result in reduction in colonization and shedding of *E. coli* in an animal. The Examiner further alleges that it is inherent that the methods of Hideo *et al.* are to be practiced in animals since Wright *et al.* teach that *E. coli* O157:H7 infects animals.

This rejection is respectfully traversed. As indicated herein, claims 53-55 are the only pending independent claims, and therefore these rejections will be addressed with respect to these claims only. The remaining claims at issue under these rejections are dependent claims and by definition subject to the limitations of claims 53, 54 or 55. Claims 53-55 are directed to methods for eliciting an immune response against an A/E pathogen or component thereof, or for reducing

colonization or shedding of an A/E pathogen, in an animal by administering an effective amount of a composition or culture supernatant including an isolated polypeptide comprising an amino acid sequence having at least 75% sequence identity to the sequence of SEQ ID NO: 24.

The Applicants reiterate that, to support a rejection under § 102, a single prior art reference must describe each and every element, either expressly or inherently, of the rejected claims (MPEP § 2131) and the prior art reference must be enabling:

“...invalidity based on anticipation requires that the assertedly anticipating disclosure enable the subject matter of the reference and thus of the patented invention without undue experimentation.” *Elan Pharmaceuticals Inc. v. Mayo Foundation for Medical Education & Research*, 346 F.3d 1051, 68 USPQ2d 1372 (Fed. Cir. 2003), hereafter “*Elan*,” emphasis added.

Hideo *et al.* do not meet these requirements, as discussed herein.

Hideo *et al.* teach nucleotide sequences from enterohemorrhagic *E. coli* O157:H7 SAKAI (referred to hereafter as the “EHEC sequences”) and assert that these sequences are not present in non-pathogenic *E. coli* K12 (see page 40, paragraph [0010], and page 71, paragraph [0014], of the enclosed English translation of Hideo *et al.*). Hideo *et al.* also teach predicted amino acid sequences based on the identified nucleotide sequences and comparison of the amino acid sequences to known sequences from various public databases using known algorithms (see page 71, paragraph [0016] of the enclosed English translation of Hideo *et al.*). Hideo *et al.* classify the predicted amino acid sequences into twelve (12) groups (see pages 71-72, paragraph [0017] of the enclosed English translation of Hideo *et al.*), as follows: 1) Proteins having unknown function etc., 2) Proteins which have unknown function, but have significant homology to that of other bacteria, 3) Proteins comprising Insertion Sequences (IS), 4) Proteins derived from phage, 5) Regulatory elements, 6) Proteins relating to fimbriae, 7) Proteins relating to transportation of substance, 8) Proteins relating to synthesis of lipopolysaccharide, 9) Proteins relating to metabolism, 10) Proteins processing DNA/RNA, 11) Proteins relating to pathogenicity, and 12) Other proteins.

Hideo *et al.* also teach that:

...a protein predicted to be a cell surface protein (membrane protein, especially, OMP, lipoprotein) in them or its gene (or nucleic-acid molecule) may be useful for production of an antibody, vaccine composition, diagnosis of O-157 infection and the like. Furthermore, there is a possibility that they include a protein which has an important function in O-157, for example, transportation and metabolism of a substance, processing of nucleic acids, and relates to a regulatory element and pathogenicity. They are to be useful for diagnosis and therapy of O-157 infection.” (see pages 267-270, paragraph [0031] of the enclosed English translation of Hideo *et al.*, emphasis added)

...

O-157 specific nucleic-acid molecule of the present invention, a gene included in it, peptide and nucleic-acid sequence encoded by the gene are useful for diagnosis and/or therapy of O-157 infection and prevention of symptom occurred by the infection. They can also be used for detection of the presence of O-157 in a sample and classification of its strain. Furthermore, they can also be used for screening of useful compounds for prevention and/or therapy of O-157 infection and symptom occurred by the infection. (see page 283, paragraph [0047] of the enclosed English translation of Hideo *et al.*, emphasis added)

...

the scope of the present invention includes a vaccine composition including genes and/or polynucleotides of the present invention, and a method for prevention and/or therapy of O-157 infection and symptom occurred by the infection. (see page 283, paragraph [0048] of the enclosed English translation of Hideo *et al.*, emphasis added)

...

The present invention relates to a peptide vaccine formulation for prevention or therapy of O-157 infection comprising effective amount of, at least one kind of, O-157 specific polypeptides having amino acid sequence set forth in the sequence lists or fragments thereof. The vaccine formulation preferably includes a pharmaceutically acceptable carrier, for example, a known adjuvant in the art. (see page 284, paragraph [0051] of the enclosed English translation of Hideo *et al.*, emphasis added).

...

The present invention relates to a method of reducing the risk of O-157 infection in patients or a method for therapy [of the infection]. This method comprises administration of the vaccine formulation of the present invention to a patient so as to reduce the risk of O-157 infection or provide therapy of infection. (see page 285, paragraph [0053] of the enclosed English translation of Hideo *et al.*, emphasis added).

The Applicants respectfully submit that these teachings of Hideo *et al.* do not rise to the level of anticipation of the claimed invention. Firstly, SEQ ID NO: 393 of Hideo *et al.*, which is asserted to be identical to SEQ ID NO: 24 of the instant application, is called out in Group 2 (Proteins which have unknown function, but have significant homology to that of other bacteria) and is described as follows at page 154 of the enclosed English translation of Hideo *et al.*:

SEQ ID NO: 393 -0.239229, 442, a minor capsid protein precursor, similar to minor capsid protein precursors for example, GpC [Bacteriophage lambda]
gi|137565|sp|P03711|VCAC#LAMBD (97% identity in 439 amino acids), capsid assembly protein containing Nu3-homolog;

The Applicants respectfully note that Hideo *et al.* do not identify SEQ ID NO: 393 as relating to pathogenicity – such sequences are listed in Group 11. The Applicants respectfully submit that capsid proteins are bacteriophage (bacterial virus) proteins used as part of the viral assembly process, and present in the viral coat upon maturation. Accordingly, a bacteriophage capsid protein would not be expected to be effective in the methods as claimed.

Secondly, Hideo *et al.* speculate that a protein that is “predicted to be a cell surface protein ... may be useful for production of an antibody, vaccine composition, diagnosis of O-157 infection ...,” that “... there is a possibility that they include a protein which has an important function in O-157, for example, transportation and metabolism of a substance, processing of nucleic acids, and relates to a regulatory element and pathogenicity...” and that such proteins “... are to be useful for diagnosis and therapy of O-157 infection.”

Accordingly, Hideo *et al.* simply raise the possibility that some of approximately two thousand (2000) sequences may be useful. This assumption appears to be based on the absence of the EHEC sequences from *E. coli* K12. Hideo *et al.* compare the sequence of the pathogenic bacterium, EHEC O157:H7, with that of the non-pathogenic K12 strain, and assert that the EHEC O157:H7 sequences that differ from the K12 sequences are pathogenic simply because EHEC O157:H7 is highly pathogenic and K12 is not. Hideo *et al.* do not provide any experimental data or other evidence to support this assertion.

The Applicants respectfully submit that the assumption that any sequence present in a pathogenic organism and absent from a non-pathogenic organism is necessarily useful is incorrect and that very few of the EHEC O157:H7-specific sequences are implicated in human disease. More specifically, non-pathogenic K12 and EHEC O157:H7 share about 80% sequence identity and are about 20% different. Given that the genomes of these organisms are about 4 million base pairs, the difference is about 800,000 base pairs. All the known virulence factors encode only about 50-100,000 base pairs (e.g., the LEE region, which encodes the Type III secretion system is 34,000 base pairs, the Shiga toxin is 1,000 base pairs, *etc.*), thus making up only a small fraction of genomic differences between non-pathogenic K12 and pathogenic EHEC O157:H7. For example, many of the non-LEE encoded proteins have no effect on virulence, and are found in O157 but not in non-pathogenic *E. coli*. Accordingly, one of ordinary skill in the art would recognize that not all of the EHEC O157:H7-specific sequences set out in Hideo *et al.* encode virulence factors. Furthermore, the inventors of the present application were the first to identify NleA polypeptide (SEQ ID NOs: 22-24) as a virulence factor. The term “virulence factor” is understood by those of skill in the art as a molecule required to cause disease, that is not normally required for viability of the micro-organism producing it in non-disease settings. It is further well known to a skilled person that, once identified, a virulence factor is useful to induce an immune response in animals, but prior to such identification there would be no reason to conclude that any protein would be useful.

Furthermore, Hideo *et al.* make it clear that the contemplated use of the disclosed sequences is in the context of treating infection in a patient. The term “infection” is defined as “[i]nvasion by and multiplication of pathogenic microorganisms in a bodily part or tissue, which may produce subsequent tissue injury and progress to overt disease through a variety of cellular or toxic mechanisms” or the “pathological state resulting from having been infected.” (see, infection, Dictionary.com, *The American Heritage® Stedman's Medical Dictionary*, Houghton Mifflin Company, <http://dictionary.reference.com/browse/infection>, accessed: November 18, 2010). Therefore, the term “infection” contemplates that the infected subject or animal exhibits symptoms of clinical disease. By contrast, ruminants may be colonized by and shed highly

virulent A/E pathogens, or exhibit an immune response against an A/E pathogen or component thereof, without ever exhibiting symptoms of overt disease.

Accordingly, the Applicants respectfully submit that Hideo *et al.* do not teach methods relating to use of an isolated polypeptide comprising an amino acid sequence having at least 75% sequence identity to the sequence of SEQ ID NO: 24 in ruminants, as claimed, and therefore do not anticipate the claimed invention.

For the sake of completeness, the Applicants note that the Examiner also asserts that Wright *et al.* teach that *E. coli* O157:H7 infects animals. With respect, Wright *et al.* disclose that *E. coli* O157:H7 “was the first EHEC strain identified in humans and remains the most common infectious cause of bloody diarrhea and hemorrhagic colitis in humans” (col. 1, ll. 34-36, emphasis added). Wright *et al.* also disclose that “...cattle, pigs, lambs and poultry may all be environmental reservoirs for verocytotoxin-producing enterohemorrhagic *E. coli*” (col. 1, ll. 49-51, emphasis added). The term “reservoir” is defined as “[a]n organism or a population that directly or indirectly transmits a pathogen while being virtually immune to its effects” (see, reservoir, Dictionary.com, *The American Heritage® Stedman's Medical Dictionary*. Houghton Mifflin Company. <http://dictionary.reference.com/browse/infection>, accessed: November 18, 2010). Accordingly, the Examiner is in error in stating that Wright *et al.* teach that *E. coli* O157:H7 infects animals since the term “infection” contemplates that the infected subject or animal exhibits symptoms of clinical disease.

Rejections Under 35 U.S.C. § 103

Claims 53-58, 71-72, 86 and 90-94 are rejected under 35 U.S.C. 103(a) as allegedly obvious over Hideo *et al.* (*supra*) in view of Wright *et al.* (*supra*).

More specifically, the Examiner alleges that Hideo *et al.* disclose a method of eliciting an immune response against *E. coli* O157:H7 by administering an effective amount of a composition

for inducing an immune response against *E. coli* O157:H7 comprising a protein identical to SEQ ID NO: 24. The Examiner further alleges that Hideo *et al.* also disclose treating an infection by *E. coli* O157:H7 using the composition. The Examiner further alleges that treatment of the *E. coli* infection will result in reduction in colonization and shedding of *E. coli* in an animal. While the Examiner concedes that Hideo *et al.* do not disclose that the animal is a ruminant or bovine or ovine or human, the Examiner alleges that it is inherent that the methods of Hideo *et al.* are to be practiced in animals since Wright *et al.* teach that *E. coli* or *E. coli* O157:H7 infect animals, such as cattle, lamb and humans and causes diarrhea.

The Examiner therefore alleges that it would have been *prima facie* obvious to one of ordinary skill in the art at the time the instant invention was made to have used the method of Hideo *et al.* for animals such as cattle, lamb and humans, thus resulting in the instant invention with a reasonable expectation of success. The Examiner finds the motivation to do so in the teachings of Wright *et al.* that *E. coli* or *E. coli* O157:H7 infect cattle, lamb and humans and cause diarrhea.

This rejection is respectfully traversed. The Applicants respectfully submit that, further to the Examination Guidelines for Determining Obviousness Under 35 U.S.C. § 103 in View of the Supreme Court Decision in KSR International Co. v. Teleflex Inc. (72 Fed. Reg. 57,526 (Oct. 10, 2007); hereafter the "Guidelines"), a proper rejection under 35 U.S.C. § 103 requires:

1. a finding that the prior art included each element claimed, although not necessarily in a single prior art reference, with the only difference between the claimed invention and the prior art being the lack of actual combination of the elements in a single prior art reference;
2. a finding that one of ordinary skill in the art could have combined the elements as claimed by known methods, and that in combination, each element merely would have performed the same function as it did separately;

3. a finding that one of ordinary skill in the art would have recognized that the results of the combination were predictable; and
4. whatever additional findings based on the *Graham* factual inquiries may be necessary, in view of the facts of the case under consideration, to explain a conclusion of obviousness.

In the present case, as discussed herein, the teachings of Hideo *et al.* are speculative and the teachings of Wright *et al.* do not cure the defect in Hideo *et al.*

More specifically, Hideo *et al.* compare the sequence of the pathogenic bacterium, EHEC O157:H7, with that of the non-pathogenic K12 strain, and speculate that the approximately 2000 EHEC O157:H7 sequences that differ from the K12 sequences are pathogenic simply because EHEC O157:H7 is highly pathogenic and K12 is not. For example, Hideo *et al.* speculate that a protein that is “predicted to be a cell surface protein ... may be useful for production of an antibody, vaccine composition, diagnosis of O-157 infection ...,” that “... there is a possibility that they include a protein which has an important function in O-157, for example, transportation and metabolism of a substance, processing of nucleic acids, and relates to a regulatory element and pathogenicity...” and that such proteins “... are to be useful for diagnosis and therapy of O-157 infection.” Hideo *et al.* do not provide any experimental data or other evidence to support these speculations.

As is discussed in detail above, the assumptions made by Hideo *et al.* are incorrect and very few of the EHEC O157:H7-specific sequences are implicated in human disease. Accordingly, one of ordinary skill in the art would recognize that not all of the EHEC O157:H7-specific sequences set out in Hideo *et al.* encode virulence factors.

Furthermore, SEQ ID NO: 393 of Hideo *et al.*, which is asserted to be identical to SEQ ID NO: 24 of the instant application, is identified as being of unknown function but similar to a capsid protein, which are bacteriophage (bacterial virus) proteins used as part of the viral

assembly process and would not be expected to be effective in the methods claimed in the instant application. As discussed above, the inventors of the present application were the first to identify NleA polypeptide (SEQ ID NOs: 22-24) as a virulence factor, which would then lead a skilled person to conclude that it would be useful to induce an immune response. Prior to such identification there would be no reason to conclude that any protein would be useful in the methods claimed in the instant application.

Finally, Hideo *et al.* make it clear that the contemplated use of the disclosed sequences is in the context of treating infection in a patient rather than in ruminants, which may be colonized by and shed highly virulent A/E pathogens, or exhibit an immune response against an A/E pathogen or component thereof, without ever exhibiting symptoms of overt disease.

Turning to Wright *et al.*, this reference discloses that *E. coli* O157:H7 “was the first EHEC strain identified in humans and remains the most common infectious cause of bloody diarrhea and hemorrhagic colitis in humans” (col. 1, ll. 34-36, emphasis added). Wright *et al.* also disclose that “...cattle, pigs, lambs and poultry may all be environmental reservoirs for verocytotoxin-producing enterohemorrhagic *E. coli*” (col. 1, ll. 49-51, emphasis added). The term “reservoir” is defined as “[a]n organism or a population that directly or indirectly transmits a pathogen while being virtually immune to its effects” (see, reservoir, Dictionary.com, *The American Heritage® Stedman's Medical Dictionary*. Houghton Mifflin Company. <http://dictionary.reference.com/browse/infection>, accessed: November 18, 2010). Therefore, contrary to the Examiner's assertion, Wright *et al.* do not teach that *E. coli* O157:H7 infects ruminants since the term “infection,” as indicated herein, contemplates that the infected subject or animal exhibits symptoms of clinical disease.

Accordingly, one of ordinary skill in the art would not have recognized that the results of the combination of Hideo *et al.* and Wright *et al.* were predictable since Hideo *et al.* provide no guidance as to which of over 2000 sequences may be useful and Wright *et al.* do not cure this defect. Therefore, Hideo *et al.*, considered alone or in combination with Wright *et al.*, do not render the claimed invention obvious.

Claims 53-55, 71-72, 86, 88-89 and 90 are rejected under 35 U.S.C. 103(a) as allegedly obvious over Hideo *et al.* (*supra*) as evidenced by Wright *et al.* (*supra*) in view of Finlay *et al.* (*supra*).

More specifically, the Office Action alleges that Hideo *et al.* disclose a method of eliciting an immune response against *E. coli* O157:H7 by administering an effective amount of a composition for inducing an immune response against *E. coli* O157:H7 comprising a protein identical to SEQ ID NO: 24. The Office Action further alleges that Hideo *et al.* also disclose treating an infection by *E. coli* O157:H7 using the composition. The Office Action further alleges that treatment of the *E. coli* infection will result in reduction in colonization and shedding of *E. coli* in an animal. The Office Action further alleges that it is inherent that the methods of Hideo *et al.* are to be practiced in animals since Wright *et al.* teach that *E. coli* or *E. coli* O157:H7 infect animals.

While the Office Action concedes that Hideo *et al.* do not disclose that the composition further comprises EspA, EspB, EspD, EspC, intimin and Tir or an adjuvant, the Office Action alleges that Finlay *et al.* teach methods for eliciting an immune response against an A/E pathogen or component thereof, or for reducing colonization of an A/E pathogen, or of reducing shedding (thus allegedly treating an infection by an A/E pathogen) in an animal by administering an effective amount of a composition comprising a culture supernatant where the composition includes EspA, EspB, EspD, EspC, intimin and Tir and/or further includes an adjuvant. The Office Action further alleges that Finlay *et al.* teach that the composition treats the EHEC infection and/or reduces colonization of the animal and teach that administration of the composition to an animal stimulates an immune response against one or more secreted antigens, such as EspA and Tir, which blocks attachment of the EHEC to intestinal epithelial cells.

The Office Action therefore alleges that it would have been *prima facie* obvious to one of ordinary skill in the art at the time the instant invention was made to have combined the composition of Hideo *et al.* with that of Finlay *et al.*, thus resulting in the instant method (wherein the composition further comprises EspA, EspB, EspD, EspC, intimin and Tir or further comprises

an adjuvant) with a reasonable expectation of success. The Office Action finds the motivation to do so because both compositions are allegedly individually taught in the prior art to be useful for the same purpose *i.e.*, inducing an immune response against *E. coli* EHEC O157:H7 and Finlay *et al.* allegedly provide additional motivation in that administration of the composition to an animal stimulates an immune response against one or more secreted antigens, such as EspA and Tir, that blocks attachment of the EHEC to intestinal epithelial cells.

This rejection is respectfully traversed. As discussed herein, in the sections addressing the rejections under 35 U.S.C. 102(b) and 103(a) with respect to Hideo *et al.*, one of ordinary skill in the art would not have recognized that the results of the combination of Hideo *et al.* and Wright *et al.*, with or without Finlay *et al.*, were predictable since Hideo *et al.* provide no guidance as to which of approximately two thousand (2000) sequences may be useful and Wright *et al.* do not cure this defect. The addition of Finlay *et al.* also do not cure the defects in Hideo *et al.* and Wright *et al.* More specifically, as indicated herein, the inventors of the present application were the first to identify NleA polypeptide (SEQ ID NOs: 22-24) as a virulence factor, which would then lead a skilled person to conclude that it would be useful to induce an immune response. Prior to such identification there would be no reason to conclude that any protein would be useful in the methods claimed in the instant application and Finlay *et al.* do not provide such identification. Accordingly Hideo *et al.*, considered alone or in combination with Wright *et al.* and/or Finlay *et al.*, do not render the claimed invention obvious.

Conclusion

The Applicants respectfully request that a timely Notice of Allowance be issued in this case.

Date: January 5, 2011

Respectfully submitted,

By: / Andrew T. Serafini/
Andrew T. Serafini, Reg. No. 41,303
Fenwick & West LLP
801 California Street
Mountain View, CA 94041
Telephone: (206) 389-4596
Facsimile: (650) 938-5200

Appendix A

Molecular Characterization and Distribution of Genes Encoding Members of the Type III Effector NleA Family among Pathogenic *Escherichia coli* Strains[†]

Kristina Creuzburg and Herbert Schmidt*

Institute of Food Science and Biotechnology, Department of Food Microbiology, University of Hohenheim, 70599 Stuttgart, Germany

Received 5 January 2007/Returned for modification 24 February 2007/Accepted 23 May 2007

In this study, we investigated the occurrence of the previously described gene *nleA*₄₇₉₅ and variants of *nleA*, putatively encoding non-locus-of-enterocyte-effacement-encoded type III effector proteins with functions that are unknown. *nleA* variants were detected in 150 out of 170 enteropathogenic *Escherichia coli* strains and enterohemorrhagic *E. coli* strains, two of them being *eae* negative. Besides the known variants *nleA*₄₇₉₅, Z6024, and the *espI*-like gene, 11 novel *nleA* variants with different lengths and sequence identities at the deduced amino acid level (between 71% and 96%) have been identified. Whereas most of the serogroups associated with more severe disease were quite homogenous with respect to the presence of a particular *nleA* variant, other serogroups were not. Moreover, Southern blot hybridization revealed that certain strains carry two copies of *nleA* in their chromosome, frequently encoding different variants. In most cases, the open reading frame of one of the copies was disrupted, usually by an insertion element. Furthermore, transmission of the type III effector-encoding gene could be shown by transduction of *nleA*-carrying bacteriophages to a laboratory *E. coli* strain.

Enterohemorrhagic *Escherichia coli* (EHEC) and enteropathogenic *E. coli* (EPEC) can cause serious gastrointestinal diseases and are able to damage the gut epithelia of their hosts by a sophisticated mechanism of attachment and effacement (11). Following adherence to intestinal cells, attaching and effacing (A/E) *E. coli* organisms interfere with cytoskeletal processes and produce a specific histopathological feature that is characterized by localized destruction of the brush border microvilli and intimate adhesion of the bacteria to the plasma membrane of the host cells (21). The development of A/E lesions is mediated by a type III secretion system (T3SS), which is able to translocate effector proteins via a needle complex directly in the cytoplasm of host cells (18). The machinery of this secretion system and its effector proteins are located within the bacterial chromosome on a pathogenicity island termed the locus of enterocyte effacement (LEE) (14).

It has been shown that additional effector proteins encoded by genes outside the LEE in cryptic or intact prophages are translocated by the LEE-encoded T3SS. The majority of these effectors have been identified by a proteomics approach with the mouse A/E pathogen *Citrobacter rodentium* (8) as well as by using bioinformatics, proteomics, and translocation assay approaches with the *E. coli* O157:H7 strain RIMD 0509952 (37). This group of non-LEE-encoded effectors also includes Cif (24), NleA/EspI (16, 27), TecP/EspF₇ (3, 15), EspI (7), NleB (20), and EspK (38). The cycle-inhibiting factor Cif blocks the cell cycle at the G₂/M-phase transition and is involved in the

formation of stress fibers (24). The Tir cytoskeleton coupling protein TecP/EspF₇ binds N-WASP and leads to Nck-independent actin polymerization (3, 15). EspI may play a role in host survival and pathogen transmission (7). NleB is probably a virulence determinant (20), whereas EspK could be involved in intestinal colonization (38).

The non-LEE-encoded effector NleA/EspI of *C. rodentium* shows 81% identity at the amino acid level to the protein Z6024, encoded by phage CP-933P in *E. coli* O157:H7 strain EDL933 (30), and 78% and 76% identity, respectively, to NleA₄₇₉₅, which is encoded by the Stx1-converting prophage BP-4795 of *E. coli* O84:H4 strain 4795/97 (6), and the EspI-like protein, encoded together with Cif by a prophage in the genome of the rabbit EPEC O103:H2 strain E22 (24). The non-LEE-encoded effectors NleA and NleA₄₇₉₅ localize close to the Golgi apparatus of HeLa cells (6, 16). Moreover, experiments with a mouse model showed that NleA/EspI is necessary for virulence (16, 27), but the function of this effector protein is still unknown. Mundy et al. (26) examined 232 EPEC and 93 EHEC strains for the presence of *espI* using colony hybridizations. They could detect *espI* in 53% of the LEE-positive EPEC strains tested. In *eae*-positive EHEC isolates, it was found more frequently. Consequently, 37 of 43 (86%) LEE-positive EHEC strains contained *espI*, and the authors assumed a correlation between the presence of *espI* and certain intimin subtypes in EPEC strains. However, it was not possible to define such an association for the occurrence of *espI* and a specific intimin type in EHEC strains. Furthermore, they could detect *espI* more commonly in strains from patients suffering from a more severe disease (26).

The aim of the present study was to determine the distribution of *nleA*₄₇₉₅ and related variants among pathogenic intestinal *E. coli* strains. Furthermore, we were interested in a

* Corresponding author. Mailing address: Department of Food Microbiology, Institute of Food Science and Biotechnology, University of Hohenheim, Garbenstraße 28, 70599 Stuttgart, Germany. Phone: 49-711/459-22305. Fax: 49-711/459-24199. E-mail: hschmidt@uni-hohenheim.de.

[†] Published ahead of print on 6 June 2007.

TABLE 1. PCR primers, cycling conditions, and PCR product lengths

Primer for <i>nleA</i> target gene variants	Primer nucleotide sequence ^a	PCR condition		Length of PCR product (bp)
		Annealing temp (°C)	Elongation time (s)	
V83-for2	5'-ACAGCAACATGACACCGGAATGC-3'	58	90	9,59–1,112
V83-rev2	5'-CTTCCATCGCAGTATATCAGC-3'			
V83-for2	5'-ACAGCAACATGACACCGGAATGC-3'	55	90	1,015–1,168
V83-rev3	5'-GATATCGATGACACCATCTCTCAGG-3'			
VarA-for ^b	5'-TATTAAGCTGTCCACATCGG-3'	50	120	1,434–1,584
VarA-rev ^b	5'-TGGTGATATTTGTTTGTGGGG-3'			
VarA-for ^c	5'-TATTAAGCTGTCCACATCGG-3'	50	120	1,333–1,483
VarA-rev2 ^a	5'-AGCTTAGACATCTGTTCTCG-3'			

^a Primers designed to amplify *nleA* and its variants. Asterisks in first column indicate primers used for cycle sequencing reactions for amplification and sequencing of the whole ORF.

possible association and correlation of the presence of *nleA* variants with serotypes and *eae* types.

MATERIALS AND METHODS

Bacterial strains. The 170 bacterial strains used in this study mainly were taken from our strain collection. A large set of strains was isolated during routine diagnostic work in the laboratory of Helge Karch at the Institute of Hygiene and Microbiology, University of Würzburg, Germany, in the years 1977 to 2001. Other strains were provided by colleagues during the European Union project QLK2-20060, and the sequences of some of these strains already have been published (1). Strains with the prefix CB originate from Lohar Beutl, Federal Institute for Risk Assessment, Berlin, Germany. Most of the E. coli O84 strains were a gift of Helmut Tschape, Robert-Koch Institute, Wernigerode, Germany, and strain S21195 was provided by Ulrich Busch, Bayerisches Landesamt für Gesundheit und Lebensmittelsicherheit, Oberschleißheim, Germany. The *E. coli* O103:H2 strain UT1 was donated by Phil I. Tarr, Washington University School of Medicine, St. Louis, MO. Other strains included in this study were H.18 (13), E2348/69 (22), RDE-1 (4), EDL933 (28), PMK5 (25), CF11201 (10), CL37 (19), and 95NR1 (39). Serotype, origin, and disease association of the strains are described in Table 2. The *E. coli* K-12 strain C600 was used as a control in different experiments, and the *E. coli* K-12 strain C600, harboring plasmid pK18 (31), was used in the transduction experiments.

Molecular techniques. Amplification of Shiga toxin genes (*stx*) and characterization of *stx*, *stx*, and *stx* were carried out by PCR as described previously (5, 23, 35, 36). Detection of *eae* and its subtypes α , β , γ , δ , and ϵ was performed as described previously (33, 40). The PCR protocols for the specific detection of other *eae* subtypes have been modified by increasing the annealing temperature to 67°C (for β and ϵ), 52°C/62°C (for γ), and 58°C (for α) (29, 34, 40).

The amplification of *nleA* variants for subsequent restriction was performed in a total volume of 70 μ l. Other PCRs were carried out in a total volume of 50 μ l containing 5 to 10 μ l of bacterial suspension and containing one to three single colonies in 150 mM NaCl, 200 μ M of each deoxynucleoside triphosphate, 30 pmol of each primer, 1 \times Taq polymerase buffer, and 3 U of Taq DNA polymerase (Genoscreen Bioscience). PCR started with an initial denaturation for 5 min at 94°C, followed by 30 cycles of amplification, consisting of denaturation for 30 s at 94°C, annealing for 60 s at a specific temperature (Table 1), and elongation at 72°C for a certain time course (Table 1). After the last cycle, a final elongation step of 5 min at 72°C completed the reaction.

Restriction of PCR products was performed, according to the manufacturer's recommendations, with BstE1 and PstI (Fermentas). For differentiation of *nleA* variants from *nleA* and *nleA*, the restriction endonucleases BstE1, NciI, and SphI were used, and the variants *nleA*, *nleA*, and the *exp*-like gene were distinguished by restriction with Bpu102I and CseI.

For DNA sequencing, *nleA* variants were amplified with primer pairs VarA-for/VarA-rev and VarA-for/VarA-rev2. To separate mixed PCR products, the respective product was extracted either from a gel using the QIAquick gel extraction kit (QIAGEN) or by being cloned into the pCR2.1-TOPO vector using the TOPO TA cloning kit (Invitrogen) according to the manufacturers' instructions. DNA sequencing was performed with the CEQ 8000 genetic analysis system (Beckman Coulter) using the CEQ dye terminator cycle sequencing quick-start kit (Beckman Coulter) by following the manufacturer's recommendations. The sequences obtained from the raw data were edited and analyzed with BioEdit (17).

Preparation of genomic DNA. For preparation of genomic DNA, 2 ml of an overnight culture was centrifuged for 10 min at 6,000 rpm, resuspended in 1 ml 0.9% NaCl solution, and centrifuged again for 5 min at 13,000 rpm in a bench-top centrifuge. The bacterial pellet was resuspended in 400 μ l STET buffer (233 mM sucrose, 50 mM Tris-HCl [pH 8.0], 20 mM EDTA [pH 8.0], 5% [vol/vol] Triton X-100, and 1.8 mg/ml lysozyme), incubated for 5 min at room temperature, heated for 1 min to 100°C, immediately cooled on ice for 2 min. After this procedure, 30 μ l of (wt/vol) sodium dodecyl sulfate and 3 μ l proteinase K (20 mg/ml; Cal Roth) were added, and the solution was incubated for 1 h at 56°C. After the addition of 3 μ l RNase A (100 mg/ml) (Sigma-Aldrich) and 200 μ l distilled water, the mixture was incubated for another 30 min at 37°C. Finally, after extraction with an equal volume of phenol-chloroform-isoamyl alcohol (25:24:1), the liquid phase was transferred into a fresh tube. The genomic DNA was precipitated with 2.5 volumes of ethanol and a 1/10 volume of sodium acetate (pH 7.2) for 30 min at –20°C, followed by centrifugation for 30 min at 13,000 rpm and 4°C. The pellet was washed with 70% ethanol and dissolved in 100 μ l distilled water.

Southern blot hybridization. Ten micrograms of genomic DNA was digested overnight with HinfI (Fermentas), separated on a 0.8% agarose gel, and transferred to a nylon membrane by vacuum blotting (32). Southern blot hybridization was performed with the DIG DNA labeling and detection kit (Roche Diagnostics) by following the manufacturer's recommendations. The hybridization was conducted at 68°C overnight using an *nleA* probe. This probe was generated from *E. coli* strain 4795/97 by PCR with the primer pair V83-for2/V83-rev3 after purification of two gel extracts with the QIAEX II gel extraction kit (QIAGEN) and was labeled with the Klenow fragment of the DNA polymerase included in the DIG DNA labeling and detection kit (Roche Diagnostics). For identification of the size of the obtained DNA fragments, a GeneRuler 1-kb DNA ladder and a λ -mix marker 19 (Fermentas) were used.

Phage transduction. An overnight culture of the respective *E. coli* strain was used to inoculate 200 ml of Luria-Bertani (LB) broth containing 1 mM CaCl₂, followed by an incubation with vigorous shaking until an optical density at 600 nm of 0.8 was obtained. After adjusting the culture with 0.05 to 0.25 μ g/ml norfloxacin (25 mg/ml in glacial acetic acid) and addition of 1 ml of SM CaCl₂, the flask with the bacterial suspension was incubated overnight. The phage particles were separated from the cell debris by centrifugation (7,500 \times g, 30 min, 4°C), followed by filtration through a funnel filter (Whatman). To remove bacterial nucleic acids, DNase I and RNase A (Sigma-Aldrich) were added to final concentrations of 0.05 μ g/ml each. After incubation at 37°C for 45 min, sodium chloride was added to a final concentration of 5.8% (wt/vol) and was dissolved, and the solution was incubated on ice for 1 h. After a centrifugation step (7,500 \times g, 10 min, 4°C), the phage particles were precipitated by adjusting the supernatant to 10% (wt/vol) polyethylene glycol 6000. After polyethylene glycol 6000 was dissolved at room temperature, the mixture was incubated on ice for 1 h. Phage particles were harvested by centrifugation (9,500 \times g, 30 min, 4°C). The resulting phage pellet was dried at room temperature and dissolved in 1 ml of SM buffer (100 mM NaCl, 8 mM MgSO₄ \cdot 7H₂O, 50 mM Tris-HCl [pH 8], and 0.01% [wt/vol] gelatin).

For transduction, 100 μ l of different dilutions of the obtained phage lysate was mixed with 100 μ l bacterial culture of *E. coli* C600/pK18 (optical density at 600 nm, ~0.8) and 2 μ l 1 M CaCl₂. The culture was incubated for 4 h at 37°C without shaking. Afterwards, 1 ml LB medium, 70 μ l 1 M CaCl₂, and 50 μ l kanamycin were added, followed by an incubation for 48 h at 37°C and shaking at 180 rpm.

TABLE 2. Distribution of *stx* types, *eae* subtypes, and *nleA* variants among 170 pathogenic *E. coli* isolates and restriction fragment lengths from Southern blot hybridizations with an *nleA*₄₇₉₅ probe

Scrotype (no. of isolates)	Host	Disease ^a (no. of cases)	<i>stx</i> type	<i>eae</i> type	<i>nleA</i> variant(s)	Hybridization product size(s) in kb (no. of products)
O15:H (1)	Rabbit	D		β	Z6024	
O26:H11 (1)	Human	NK	1	β	Z6024 ^b <i>nleA8-1</i>	8.6, 11.8
O26:H11 (3)	Human	HUS	2	β	Z6024 ^b <i>nleA8-1</i>	12, 24 (2)
O26:H11 (2)	Human	HUS (1), NK (1)	2	β	<i>nleA8-1</i>	25
O26:H11 (2)	Human	HUS (1), D (1)	1	β	Z6024 ^b <i>nleA8-1</i>	12, 24
O26:H ⁺ (1)	Human	NK	1/2	β	Z6024 ^b <i>nleA8-1</i>	8.6, 11.8
O26:H ⁺ (2)	Human	D (1), AS (1)	1	β	Z6024 ^b <i>nleA8-1</i>	12, 24
O26:H ⁺ (2)	Human	NK (1), AS (1)	2	β	Z6024 ^b <i>nleA8-1</i>	12, 24
O26:H ⁺ (5)	Human	HUS (4), D (1)	2	β	Z6024 ^b <i>nleA8-1</i>	12, 24
O49:H2 (3)	Human	D		β	<i>espI</i> -like gene	20 (2)
O49:H2 (1)	Human	D		β		No signal
O49:H2 (1)	Rabbit	NK		β	<i>espI</i> -like gene	4.3
O49:H10 (1)	Dog	NK		κ	Z6024	9
O49:H18 (1)	Pig	NK		κ	<i>nleA2</i>	16
O49:H35 (1)	Cattle	NK		κ	<i>nleA10</i>	17
O49:NM (1)	Human	HC		κ		4.8
O49:NM (1)	Cattle	NK		κ		No signal
O49:NM (1)	Dog	D				No signal
O49:NM (3)	Pig	NK		β	<i>nleA8-1</i> , <i>nleA8-1</i>	4.5, 23
O84:H2 (1)	Cattle	NK	1	ζ	<i>nleA4</i>	24
O84:H2 (1)	Human	D	1	ζ	<i>nleA9</i> , <i>nleA8-2^c</i>	7, 30
O84:H2 (1)	Human	NK	1	ζ	<i>nleA9</i> , <i>nleA8-2^c</i>	30
O84:H2 (1)	Human	D	1	ζ	<i>nleA7</i> , <i>nleA8-2^c</i>	6.8, 30
O84:H4 (1)	Human	D	1	ζ	<i>nleA4795</i>	11
O84:H28 (2)	Human	NK	1	ζ	<i>nleA4795</i>	11 (1)
O84:H ⁺ (1)	Human	D	1	ζ	<i>nleA9</i> , <i>nleA8-2^c</i>	30
O84:H ⁺ (1)	Human	D	1	ζ	<i>nleA7</i> , <i>nleA8-2^c</i>	6.8, 30
O84:H ⁺ (2)	Cattle/sheep	NK	1	ζ	<i>nleA7</i> , <i>nleA8-2^c</i>	6.8, 30
O84:H ⁺ (1)	Pig	NK	0		<i>nleA8-2</i> , <i>nleA8-1^b</i>	13, 25
O84:nt (1)	Human	D	1	ζ	<i>nleA9</i>	30
O84:nt (2)	Human	D (1), NK (1)	1	ζ	<i>nleA7</i> , <i>nleA8-2^c</i>	6.8, 30
O84:nt (1)	Human	D	1	ζ	<i>nleA9</i> , <i>nleA8-2^c</i>	7, 30
O103:H2 (2)	Rabbit/chicken	D		β	<i>espI</i> -like	13
O103:H2 (1)	Meat		1	ε	<i>espI</i> -like	24.5
O103:H2 (1)	Human	SIDS		ε	<i>espI</i> -like	24.5
O103:H2 (1)	Human	D		β	<i>espI</i> -like, <i>nleA11</i>	7, 24.5
O103:H2 (3)	Human	HUS	2	ε	<i>espI</i> -like	24.5
O103:H2 (3)	Human	NK (2), AS (1)	1	ε	<i>espI</i> -like	24.5 (3)
O103:H2 (1)	Human	NK	1		<i>espI</i> -like	24.5
O103:H2 (4)	Human	HUS (3), UTI (1)	1	ε	<i>espI</i> -like	24.5 (3)
O103:H17 (1)	Dog	HC				No signal
O103:H11 (1)	Human	NK	1	β	<i>nleA8-1</i>	25
O103:H18 (1)	Human	D	1	ε	<i>espI</i> -like	24.5
O103:H ⁺ (2)	Human	HUS (1), D (1)	1	ε	<i>espI</i> -like	24.5 (1)
O103:H ⁺ (1)	Human	HUS	2	ε	<i>espI</i> -like	24.5
O111:H2 (1)	Human	NK		β	<i>espI</i> -like	25
O111:H2 (1)	Human	NK		θ	Z6024	11.2
O111:H2 (2)	Human	D	1	θ	<i>nleA8-1</i>	27 (1)
O111:H8 (1)	Human	HUS	1	θ	<i>nleA8-1</i>	27
O111:H ⁺ (1)	Human	NK		β	<i>espI</i> -like	
O111:H ⁺ (6)	Human	HUS (2), D (4)	1	θ	<i>nleA8-1</i>	27 (4)
O111:H ⁺ (7)	Human	HUS	1/2	θ	<i>nleA8-1</i>	27 (2)
O111:H ⁺ (1)	Human	HUS	1/2		<i>nleA8-1</i>	27
O118:H5 (1)	Human	NK		κ	Z6024	
O125:H ⁺ (1)	Human	D		η	<i>nleA8-2</i>	12
O127:H6 (1)	Human	NK		α		No signal
O128:H2 (3)	Human	D (2), AS (1)	1/2			No signal (3)
O128:H2 (7)	Human	D		β	<i>espI</i> -like	20 (3)
O128:H2 (2)	Rabbit/chicken	D (1), AS (1)		β	<i>espI</i> -like	20
O128:H2 (1)	Sheep	AS	1/2			No signal
O128:H2 (2)	Pigeon	AS	2f	β	<i>espI</i> -like	10 (1)
O128:B12 (1)	Human	D	2f	β	<i>espI</i> -like	10
O128:H ⁺ (1)	Human	D	1			No signal
O128:H ⁺ (2)	Human	D (1), AS (1)	1/2			No signal (2)
O145:H4 (1)	Human	NK		ι		No signal

Continued on facing page

TABLE 2—Continued

Serotype (no. of isolates)	Host	Dissease ^a (no. of cases)	<i>stx</i> type	<i>eae</i> type	<i>nleA</i> variant(s)	Hybridization product size(s) in kb (no. of products)
O145:H28 (1)	Human	HUS	2	γ	<i>nleA3</i>	5
O145:H28 (2)	Human	D	1	γ	<i>nleA5</i>	5.5
O145:H28 (1)	Human	HUS	2	γ	<i>nleA6-1</i>	9.8
O145:H28 (1)	Human	D				No signal
O145:H28 (1)	Pig	AS		γ	<i>nleA11</i>	19
O145:H34 (1)	Human	D		γ		No signal
O145:NM (6)	Human	HUS (5), D (1)	2	γ	<i>nleA2</i>	30 (2)
O145:NM (2)	Human	HUS (1), D (1)	1	γ	<i>nleA2</i>	30 (2)
O145:NM (1)	Cattle	AS	1	γ	<i>nleA2</i>	30
O145:NM (1)	Pig	NK		γ	<i>nleA2</i>	
O145:NM (1)	Human	HUS	2	β	Z6024	31
O145:NM (1)	Human	HUS	1/2	γ	<i>nleA2</i>	30
O156:H11 (1)	Human	AS		ζ	1	6
O156:H18 (4)	Human	NK (3), AS (1)		θ		No signal (4)
O156:H18 (2)	Pig	NK (1), D (1)		θ	Z6024	9 (1)
O156:H21 (1)	Human	D	1/2			No signal
O156:H21 (1)	Cattle	NK	1	ζ	<i>nleA3</i> , <i>nleA8-2^c</i>	4.8, 6.2
O156:H25 (1)	Human	AS	1	ζ	<i>nleA3</i> , <i>nleA8-2^c</i>	4.8, 6.2
O156:H25 (1)	Human	NK		θ	<i>nleA6-2</i> , <i>nleA8-2^c</i>	6.2, 9.8
O156:H25 (1)	Human	NK		ζ	<i>nleA6-1</i> , <i>nleA8-2^c</i>	6.2, 9.8
O156:H25 (1)	Cattle	NK		ζ	<i>nleA3</i> , <i>nleA8-2^c</i>	5.5, 6.2
O156:H25 (4)	Sheep	NK	1	ζ	<i>nleA6-1</i> , <i>nleA8-2^c</i>	6.2, 11.5 (4)
O157:H17 (1)	Human	D		γ	Z6024	10
O157:H17 (1)	Ground beef		1/2	γ	Z6024	12
O157:H17 (4)	Human	HUS (3), AS (1)	1/2	γ	Z6024	12 (1)
O157:H17 (4)	Human	HUS (2), D (2)	2	γ	Z6024	12 (2)
O157:H17 (1)	Human	NK		γ	Z6024	
O157:H17 (2)	Human	D (1), AS (1)	1/2	γ	Z6024	12 (2)
O157:H17 (1)	Human	F	1/2	γ	Z6024	
O157:H17 (2)	Human	HUS	2	γ	Z6024	
O157:H17 (4)	Human	HUS (3), AS (1)	2	γ	Z6024	9 (1), 12 (1)

^a NK, not known; D, diarrhea; AS, asymptomatic; HUS, hemorrhagic colitis; SIDS, sudden infant death syndrome; UTI, urinary tract infection; E, enteritis.

^b Variant gene with a 1-bp deletion of either the Z6024 or *nleA8-1* sequence.

^c The 5' end of *nleA8-2* was missing from this gene.

The culture then was centrifuged for 30 min at 3,500 × g and at 4°C, and the pellet was plated on LB agar containing 50 µg/ml kanamycin.

Nucleotide sequence accession numbers. The coding sequences of the variants *nleA1* to *nleA11* have been deposited in the GenBank database under continuous accession numbers AM421995 to AM422007.

RESULTS

Distribution of *nleA* variants in EHEC and EPEC strains. One hundred seventy EHEC and EPEC strains, which were predominantly human stool isolates of serogroups O26, O103, O111, O145, and O157 and were associated with severe human disease, as well as animal and human isolates of serogroups O49, O84, O128, and O156, with minor roles in human pathogenicity, were selected for this study. The 135 human isolates originated from asymptomatic carriers and patients with symptoms ranging from diarrhea to hemolytic-uremic syndrome (HUS), which were chosen to assess the occurrence of associations with more severe disease as postulated by Mundy et al. (26). Another 33 strains originated from feces of animals, and 2 were food isolates. The strains mainly were isolated in Germany, other European countries, the United States, Canada, Brazil, and Australia. Mundy et al. (26) detected *nleA* only in association with the intimin-encoding gene *eae*. Therefore, we included 157 *eae*-positive strains. We also used 13 *eae*-negative strains.

The presence of *nleA* was determined by PCR with primer

V83-for2 in combination with either V83-rev2 or V83-rev3, which is complementary to conserved regions of this gene. Primer V83-rev3 was constructed because it was not possible to amplify a PCR product from all strains with the primer V83-rev2. Total DNA of PCR-negative strains was hybridized with an *nleA*₄₇₉₅ probe to exclude the possibility that negative PCR results were due to variations in primer-binding sites. Only one PCR-negative rabbit O49:H2 strain was detected with the *nleA*₄₇₉₅ probe. Members of the *nleA* gene family were detected in 150 out of 170 strains, and 148 of these isolates carried one of the *eae* alleles β, ε, γ, η, κ, θ, and ζ (Table 2). Interestingly, we were able to detect *nleA* in two *eae*-negative human isolates with serotypes O103:H2 and O111:H7. *nleA* genes were absent in two *eae*-positive human strains of serogroup O145. Moreover, *nleA* was not detectable in the EPEC strain I:2348/69, which expresses intimin α, in two *eae*-positive strains of serogroup O49, and in three human *eae*-positive O156:H8 isolates (Table 2). *nleA* members were present in the two food isolates but were absent from four *eae*-negative animal isolates and one *eae*-positive animal isolate. Moreover, they were found in 119 out of the 135 human isolates, whereas the 15 *nleA*-negative isolates included 8 *eae*-positive and 8 *eae*-negative *E. coli* strains. In addition, we were able to detect *nleA* genes in all 48 HUS isolates as well as in 40 out of 48 human strains associated with diarrheal disease (Table 2).

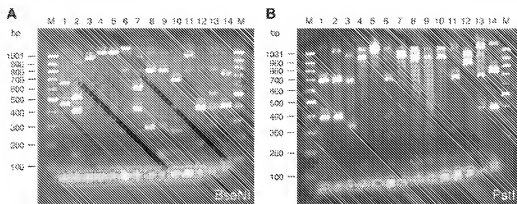


FIG. 1. Agarose gel electrophoresis of selected PCR products of *nleA* variants, each of them separately restricted with BseNI (A) or PstI (B). The PCR products were amplified from the following *E. coli* strains (the variant harbored is named in parentheses): lane 1, 4795/97 (*nleA*₄₇₉₅); lane 2, EDL933 (*Z6024*); lane 3, E22 (the *espI*-like gene); lane 4, 5721/96 (*nleA*₅₇₂₁); lane 5, 0917/99 (*nleA*₅₇₂₁); lane 6, PT272 (*nleA*₆); lane 7, CB6116 (*nleA*₇); lane 8, 3439/00 (*nleA*₈); lane 9, CF11201 (*nleA*₈); lane 10, CB6389 (*nleA*₉); lane 11, CB8745 (*nleA*₆); lane 12, CB7690 (*nleA*₁₀); lane 13, 1247/95 (*Z6024* and *nleA*₈); lane 14, DG264/4 (*nleA*₁₁). The molecular weight marker (M) is a GeneRuler 100-bp DNA ladder (Fermentas).

Restriction analyses of the *nleA* variants. The PCR products obtained with primer pairs V83-for2/V83-rev2 and V83-for2/V83-rev3 did not have the same length. Their sizes ranged from 959 bp to 1,168 bp (Table 1). Two restriction enzymes were chosen for differentiation because of the similarity of *nleA*₄₇₉₅ and *Z6024*. Separate restrictions of all PCR products with BseNI (Fig. 1A) and PstI (Fig. 1B) showed 11 different patterns among the 150 PCR products, in addition to the already known variants *nleA*₄₇₉₅, *Z6024*, and the *espI*-like gene of *E. coli*. The expected restriction pattern for *nleA/espI* of *C. rodentium* was not observed. After restriction with BseNI, PCR products obtained with strains 5721/96 and 0917/99, depicted in lanes 4 and 5, respectively, showed the same pattern (Fig. 1), whereas different patterns were obtained by using PstI. Furthermore, PCR products of strains 5721/96, 3439/00, CF11201, and CB6389 in lanes 4, 8, 9, and 10 as well as those of strains PT272 and CB8745 in lanes 6 and 11 share the same PstI restriction pattern (Fig. 1), in contrast to their varying BseNI patterns. Moreover, the addition of molecular weights of BseNI and PstI restriction fragments of strain 1247/95 in lane 13 revealed molecular weights that were approximately double the weights of the other strains. Based on this observation and the arrangement of the restriction fragments, we concluded that this PCR product is a mixture of the two *nleA* variants shown in lane 2 and lane 8 (Fig. 1).

Molecular characterization of the *nleA* variants. In order to prove the assumption that each restriction pattern represents an independent *nleA* variant, for each pattern, one strain of each serogroup was selected and the respective PCR product was sequenced. Problems arose during sequencing because of the occurrence of mixed products. This finding, along with the detection of the restriction pattern shown in Fig. 1, lane 13, led to the verification of the hypothesis that some strains may carry more than one copy of *nleA*. Therefore, Southern blot hybridization was performed (Fig. 2).

Two copies of *nleA* variants were detected in most of the *nleA*-positive isolates of serogroups O26, O84, and O156. Furthermore, two copies were detectable in three O49:NM strains, originating from pigs that probably were from the same farm,

and one human O103:H2 isolate. All the other *nleA*-positive O49 and O103 strains examined possessed only one copy of the gene. This was also true for all *E. coli* O111, O128, O145, and O157 strains (Fig. 2; Table 2).

DNA fragments were amplified by using primer VarA-for either in combination with VarA-rev, which binds in the region downstream of *nleA*, or in combination with VarA-rev2, which binds at the 3' end of the gene. For DNA sequencing of PCR products of isolates with two copies of *nleA*, the following strategies were used. If the PCR product appeared as a single band, it could be cloned directly into the pCR2.1-TOPO vector of the TOPO TA cloning kit (Invitrogen). If two PCR products were amplified, the PCR product with the expected size was extracted from a gel prior to sequencing.

DNA sequencing resulted in the identification of 11 new *nleA* variants, termed *nleA*₁ to *nleA*₁₁, besides the three known variants of pathogenic *E. coli* (Fig. 3). We defined an open reading frame (ORF) with a cutoff value of less than 97% sequence identity at the deduced amino acid level as an individual variant of *nleA*. By sequencing, two *nleA* variants were identified that shared a restriction pattern that was the same as that of the *espI*-like gene. To distinguish variants *nleA*₁, *nleA*₂, and the *espI*-like gene, PCR products with the primer pair V83-for2/V83-rev2 from all concerned strains were restricted either by Bpu1102I or by CseI. Moreover, many members of variant *nleA*₈-1 (see below) differ in 1 bp in the recognition site of BseNI, resulting in different restriction patterns. This is shown in lanes 4 and 8 in Fig. 1. To distinguish this variant, *nleA*₈-1, from *nleA*₃ and *nleA*₄, which had the restriction pattern shown in lane 4 (Fig. 1), the restriction enzymes BclI, NheI, and SphI were used.

The 15 variants, including *nleA* of *C. rodentium*, revealed sequence identities to each other of between 71% and 96% at the deduced amino acid level. Several variants showed strain-specific differences in the amino acid sequences that were caused by one to three point mutations. These mutations were not taken into consideration in this study. Only the strain-specific differences in the sequence of variant *nleA*₆ and *nleA*₈ were separated by an additional numerical suffix, be-

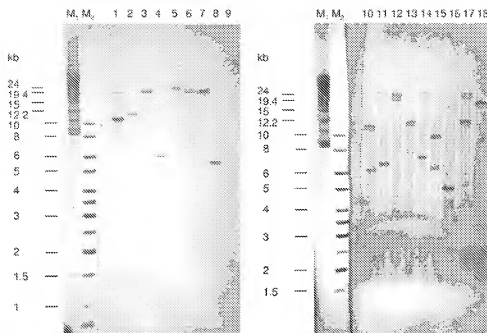


FIG. 2. Southern blot hybridization of different *nleA* variants with an *nleA*₄₇₉₅ probe. The following *E. coli* strains were used (the variant harbored is named in parentheses): lane 1, O84:H4 strain 4795/97 (*nleA*₄₇₉₅); lane 2, O157:H7 strain 0136/96 (Z6024); lane 3, O103:H2 strain 2576/97 (the *espI*-like gene); lane 4, O156:H11 strain LITEC94460 (*nleA*1); lane 5, O145:NM strain 4672/99 (*nleA*2); lane 6, O111:H2 strain 5721/96 (*nleA*8-1); lane 7, O84:H2 strain CB7197 (*nleA*4); lane 8, O145:H28 strain 0917/99 (*nleA*5); lane 9, O127:H6 strain E2348/69 (negative control); lane 10, O156:H25 strain PT272 (*nleA*6-1 and *nleA*8-2); lane 11, O84:H1 strain CB6116 (*nleA*7 and *nleA*8-2); lane 12, O26:H11 strain 3439/00 (*nleA*8-1); lane 13, O125:H1 strain CF11201 (*nleA*8-2); lane 14, O84:NM strain CB6389 (*nleA*8-2 and *nleA*9); lane 15, O156:H25 strain CB8745 (*nleA*6-2 and *nleA*8-2); lane 16, O49:NM strain CB7690 (*nleA*10); lane 17, O26:H1 strain 5720/96 (Z6024 and *nleA*8-1); lane 18, O145:H28 strain DG264/4 (*nleA*11). M₁ is the lambda mix marker; 19₂ is a GeneRuler 1-kb DNA ladder (Fermentas).

cause NleA6-2 possessed an insertion of four amino acids that were absent from NleA6-1, resulting in a deduced protein length of 462 amino acids. On the other hand, NleA8-1 and NleA8-2 varied in 10 amino acids of the C-terminal end (Fig. 3). This difference seems to be associated with specific serogroups (Table 2). Moreover, the length of the 15 deduced proteins varied between 412 and 462 amino acids (Fig. 3). These variations are due to the repeated occurrence of amino acids, in particular of alanine, serine, and threonine, in the middle region of the deduced proteins (Fig. 3). Furthermore, this region includes a putative transmembrane helix. Because of the absence of this region, this putative helix is missing from the variants *Espl*-like protein, NleA1, NleA2, NleA7, NleA10, and NleA of *C. rodentium* (Fig. 3). Another putative transmembrane helix is located in the C-terminal third of the deduced amino acid sequence. This one is present in all variants described (Fig. 3).

Most of the strains of serogroups O26, O84, and O156, as well as three O49:NM isolates from pigs, probably originating from the same farm, and one O103:H2 strain, harbored two copies of the gene *nleA* (Table 2). Moreover, one copy of *nleA* was disrupted in isolates of serogroups O26, O84, and O156. Sequencing of variant Z6024 of serogroup O26 revealed the deletion of 1 bp, resulting in a truncated, possibly nonfunctional putative protein. This was also the case for *nleA*8-1 of the O84:H1 isolate that harbored variants *nleA*8-1 and *nleA*8-2. Other strains of serogroup O84 harboring *nleA*7 and *nleA*8-2 or *nleA*9 and *nleA*8-2 showed a disrupted ORF of

variant *nleA*8-2 because of the insertion of the insertion sequence (IS) element ISEc8 651 bp upstream of the 3' end of the gene. In contrast, 117 bp of the 5' end of variant *nleA*8-2, present in serogroup O156, was missing. The inserted sequence resembled the region upstream of *nleA*₄₇₉₅ of the prophage BP-4795 and, to some extent, an ISEc8 element. Therefore, 179 bp that was in BP-4795 was missing from this sequence. Furthermore, the first 34 bp of the *espI*-like gene of one O128:H2 pigeon isolate was deleted due to insertion of an IS element, in contrast to other strains of serotype O128:H2 that harbored a complete ORF of the *espI*-like gene. Variant *nleA*4 also may encode a truncated protein because of the insertion of 5 bp located 78 bp downstream of the 5' end of the gene.

Most of the isolates of serogroups associated with severe human disease harbored variant Z6024, *nleA*8-1, or the *espI*-like gene (Table 2). Moreover, only one or three different variants of *nleA* could be detected from the serogroups O26, O157, O103, and O111. In contrast, serogroup O145 appeared to be heterogeneous, with six different variants. Whereas serotype O145:NM was nearly uniform, four variants of *nleA* were detected in the six O145:H28 isolates examined. This also was the case for serogroups O49, O84, and O156. These serogroups harbored five to six different variants, whereas *nleA*-positive O128 strains possessed only the *espI*-like gene. Variants Z6024, *nleA*8-1, and the *espI*-like gene occurred most frequently, followed by variants *nleA*8-2 and *nleA*2. In contrast, the other nine variants were detected only in one to six isolates (Table 2). Furthermore, most of the *eae* subtypes β, γ, δ, ζ, and

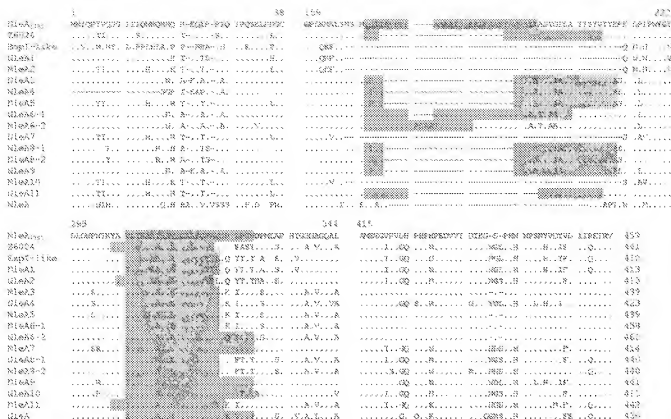


FIG. 3. Alignment of the deduced amino acid sequences of the four known *nleA* variants NleA₁₇₉₅, Z6024, the EspI-like protein, and NleA, as well as the newly discovered variants NleA1 to NleA11. Only the N-terminal regions, the two putative transmembrane helices, and the C-terminal regions are shown. The position numbers refer to the sequence of NleA₁₇₉₅. Identical amino acids are depicted by dots, and amino acids that are absent from a particular sequence are indicated by dashes. The putative transmembrane helices are labeled in gray.

e, detected in a larger number of isolates, were associated with four to eight different *nleA* variants (Table 2). Because of the rare occurrence of most of the *nleA* variants as well as the association of the *espI*-like gene, *nleA*-I, and Z6024 with at least two different *eae* subtypes, a direct correlation between a certain variant of *nleA* and a specific *eae* subtype could not be defined.

Transduction of *nleA* variants. Previously, it was shown that *nleA*₁₇₉₅ and the *espI*-like gene are located in the genome of prophages, which are fully inducible to produce phage particle progeny, whereas Z6024 is located on a cryptic prophage (6, 24, 30). In order to determine whether the newly discovered variants are within intact phages that are able to spread the *nleA*-encoded T3SS effector by horizontal gene transfer, transduction experiments were carried out. A collection of 24 pathogenic *E. coli* strains harboring different *nleA* variants were chosen as donor strains for these experiments (Table 3). *E. coli* C600/pK18 was used as the recipient to prevent the growth of single wild-type cells, which were not detached during preparation of phage lysates. The 11 isolates carrying *stx*₁ or *stx*₂ that were examined carried inducible *Stx2*-converting prophages, which were transducible in *E. coli* C600/pK18 (Table 3). All strains harboring *stx*₂ belonged to serogroups associated with severe human disease, and most of them originated from patients with HUS. Only one strain, the O145:H28 isolate C34973, also possessed the variant *nleA*-I in the genome of an inducible prophage. Furthermore, the three isolates 4795/

97, 01-08612, and CB6389 of serogroup O84, as well as the O49:NM strain CB7690, obviously harbored functional prophages carrying an *nleA* variant, which could be transduced in the *E. coli* K-12 strain C600/pK18 (data not shown). Strains CB7690, 4795/97, and 01-08612 carried variant *nleA*10 or *nleA*₁₇₉₅ in the genome of an inducible bacteriophage. Strain CB6389 possessed an intact prophage harboring variant *nleA*9 that is not disrupted by an IS element, and the similar variant *nleA*-8-2 was located in the genome of this isolate as well. Moreover, each of these three O84 strains exhibited an intact *Stx1*-converting prophage. This also was the case for the O84:H2 isolate CB7197. However, it was not possible to transduce the variant *nleA*4 of this isolate in the *E. coli* strain C600/pK18. No bacteriophages harboring *stx*₁ or a variant of *nleA* and originating from the other eight *stx*₁-positive or *stx*₂-negative strains were detectable after transduction in C600/pK18. Thus, we were able to demonstrate the transduction of five *nleA*-carrying phages with the method described above. However, we cannot exclude completely the possibility that the other 19 analyzed *E. coli* strains also harbor inducible *nleA* phages that could not be detected in this assay.

DISCUSSION

The detection of different variants of the gene *nleA* in 150 out of 170 *E. coli* strains examined shows the widespread oc-

TABLE 3. *E. coli* donor strains used for transduction of phages to the recipient *E. coli* C600/pK18 and transduced genetic traits

Strain	Serotype	Host	Disease ^a	stx type	eae type	nleA type	Transduced trait(s)
1530/99	O26:H11	Human	HUS	2	β	Z6024 ^b , nleA-1	stx ₂
3439/00	O26:H11	Human	NK	2	β	nleA-1	stx ₂
CB7690	O49:NM	Human	HC	κ	κ	nleA10	nleA10
CB7197	O84:H2	Catfish	NK	1	ζ	nleA4	stx ₃
03-06145	O84:H2	Human	D	1	ζ	nleA7, nleA8-2 ^c	stx ₂
4795/97	O84:H4	Human	D	1	ζ	nleA ₁₂₀₅	stx ₃ , nleA ₁₂₀₅
01-08612	O84:H28	Human	NK	1	ζ	nleA ₁₂₀₅	stx ₃ , nleA ₁₂₀₅
CB8966	O84:H ⁻	Pig	NK	0	θ	nleA8-2, nleA8-1 ^b	stx ₂
CB6389	O84:nt	Human	D	1	ζ	nleA9, nleA8-2 ^c	stx ₃ , nleA9
UTI	O103:H2	Human	UTI	1	ε	espI-like	stx ₂
2636/97	O103:H ⁻	Human	HUS	2	ε	espI-like	stx ₂
1639/77	O111:H ⁻	Human	D	1	θ	nleA-1	stx ₂
1187/00	O111:H	Human	HUS	1/2	γ	nleA8-1	stx ₂
14/97	O128:H2	Pigeon	AS	2f	β	espI-like ^d	stx ₂₁
0917/99	O145:H28	Human	D	1	γ	nleA5	stx ₂
4557/99	O145:H28	Human	HUS	2	γ	nleA3	stx ₂
CB4973	O145:H28	Human	HUS	2	γ	nleA6-1	stx ₂ , nleA6-1
DG264/4	O145:H28	Pig	AS	γ	γ	nleA11	stx ₂
4392/97	O145:NM	Human	HUS	2	β	Z6024	stx ₂
4672/99	O145:NM	Human	HUS	1/2	γ	nleA2	stx ₂
CB8104	O145:NM	Human	HUS	2	γ	nleA2	stx ₂
LTEC94460	O156:H1	Human	AS	ζ	ζ	nleA1	stx ₂
PT272	O156:H25	Sheep	NK	1	ζ	nleA6-1, nleA8-2 ^c	stx ₂
2492/00	O157:H ⁻	Human	HUS	2	γ	Z6024	stx ₂

^a NK, not known; HC, hemorrhagic colitis; D, diarrhea; UTI, urinary tract infection; AS, asymptomatic.^b Variant gene with a 1-bp deletion of either the Z6024 or nleA-1 sequence.^c The 5' end of nleA8-2 or the espI-like gene was missing from this gene.

currence of this non-LEE-encoded T3SS effector among pathogenic *E. coli* strains. With the exception of two strains, we could confirm the appearance of *nleA* in association with *eae* as determined by Mundy et al. (26). It has yet to be proven whether the two *eae*-negative isolates are able to secrete NleA or if the gene represents a relic of extensive genetic rearrangement without any known function. Moreover, although the function of the virulence determinant NleA is unknown, the widespread distribution of the encoding gene points to an apparent selective advantage for *E. coli* strains harboring copies of this gene.

DNA sequencing revealed 15 gene variants of different lengths. Interestingly, a region of 4 to 51 deduced amino acids approximately located in the middle of the encoded deduced protein was lacking in some variants. This region, characterized by the predominant occurrence of alanine, serine, and threonine, includes a putative transmembrane helix. Therefore, in variants of *nleA* with a deletion of 36 to 51 amino acids, this helix is missing. Thus, these variants exhibit only one putative transmembrane helix, whereas other members of the *nleA* gene family possess two helices. At present, the role of the number of helices is unknown. Possibly, there are differences in the location or function of the deduced proteins inside eukaryotic cells.

The intimin subtypes seem to be responsible for different host tissue tropisms in the intestine (9). Therefore, they may contribute to the severity of symptoms during a disease caused by EHEC or EPEC. Such an association also was shown for Stx variants. Stx2 often causes more severe disease than those caused by Stx1, whereas differences appear among the heterogeneous members of the Stx2 group (2, 12). Until now, no significant correlation could be determined between the occurrence of a certain *nleA* variant and the appearance of a specific

eae type, stx type, or pathotype. The distribution of different members of the *nleA* gene family seems to be associated more closely with the serogroup of *E. coli* strains. Thus, with the exception of *E. coli* O145 strains, the serogroups associated with more severe disease in humans predominantly contain one or two different *nleA* variants, whereas less important serogroups contain a larger number of variants (Table 2). The majority of O26, O103, O111, and O157 strains harbor Z6024, the *espI*-like gene, or *nleA*-1. On the other hand, strains of serogroups O49, O84, and O156 harbor a variety of members of the *nleA* gene family. Therefore, these strains may be depicting a pool for genetic rearrangements.

Whereas Z6024 is harbored by the cryptic prophage CP-933P (30), *nleA*₁₂₀₅ and the *espI*-like gene are carried by inducible bacteriophages (6, 24). Transduction experiments also revealed the location of *nleA*6-1, *nleA*9, and *nleA*10 in the genome of inducible phages. T3SS effector protein-encoding genes often are present at one end of bacteriophages, presumably a result of incorrect excision during the lytic life cycle. Moreover, the transduction of bacteriophages that carry a variant of the gene *nleA* to a laboratory *E. coli* strain raises the possibility that NleA-converting bacteriophages can be spread by horizontal gene transfer. Some strains harbored two copies of mostly different *nleA* variants. In the majority of the concerned isolates, one of these copies was deleted by an IS element, or some base pairs were missing. This perhaps results from the ability of bacteriophages harboring a member of the *nleA* gene family to infect the *E. coli* strain when parts of another lysogenic phage in the genome of this *E. coli* strain were destroyed by genetic rearrangements. These results indicate a major role for bacteriophages in the distribution of the members of the *nleA* gene family.

Recently, it was shown that *nleA*₄₇₉₈ is harbored together with *stx*₂ on a single bacteriophage (6). Whether this is the case for other bacteriophages, such as BP-01-08612, BP-CB6389, and BP-CB4973, remains to be elucidated.

Interestingly, the *nleA* variants investigated in this study and those investigated by other authors are linked to phage DNA. This fits with the concept that the specific characteristics of pathogenic *E. coli* strains are located on mobile genetic elements. To elucidate the role of the phage-encoded type III effectors in more detail, further research is needed.

ACKNOWLEDGMENTS

We thank Markus Kranz for skilful technical assistance. This work was supported by grant SchM1366/1-3 from the Deutsche Forschungsgemeinschaft.

REFERENCES

- Beutin, L., S. Kauffuss, S. Herold, E. Oswald, and H. Schmidt. 2005. Genetic analysis of enteropathogenic and enterohemorrhagic *Escherichia coli* serogroup O103 strains by molecular typing of virulence and housekeeping genes and pulsed-field gel electrophoresis. *J. Clin. Microbiol.* 43:1552–1563.
- Boerlin, P., S. A. McElwain, F. Boerlin-Peterson, J. B. Wilson, R. P. Johnson, and C. L. Gyles. 1999. Associations between virulence factors of Shiga toxin-producing *Escherichia coli* and disease in humans. *J. Clin. Microbiol.* 37:497–503.
- Campellone, K. G., D. Robbins, and J. M. Leong. 2004. EspF₂ is a translocated EHEC effector that interacts with Tir and N-WASP and promotes Ntk-independent actin assembly. *Dev. Cell* 7:217–228.
- Cantos, J. R., and R. K. Blake. 1977. Diarrhea due to *Escherichia coli* in the rabbit: a novel mechanism. *J. Infect. Dis.* 135:454–462.
- Cebula, T. A., W. L. Payne, and P. Feng. 1993. Simultaneous identification of strains of *Escherichia coli* serotype O157:H7 and their Shiga-like toxin by mismatch amplification mutation assay-multiplex PCR. *J. Clin. Microbiol.* 32:248–250.
- Cruzhurg, R., J. Rockenthauser, V. Kuhl, S. Herold, M. Hensel, and H. Schmidt. 2005. The Shiga toxin 1-converting bacteriophage BP-4795 encodes an NleA-like type III effector protein. *J. Bacteriol.* 187:8494–8498.
- Dahan, S., S. Wiles, R. M. La Regione, A. Best, M. J. Woodward, M. P. Stevens, R. K. Shaw, Y. Chong, S. Knutton, A. Phillips, and G. Frankel. 2005. EspI is a prophage encoded type III effector protein of attaching and effacing pathogen that modulates infection dynamics. *Infect. Immun.* 73:679–686.
- Deng, W., J. L. Puente, S. Gruenheid, Y. Li, B. A. Vallance, A. Vazquez, J. Barba, J. A. Ibarra, P. O'Donnell, P. Metanikov, K. Ashman, S. Lee, D. Goode, T. Pawson, and B. H. Finlay. 2004. Discerning virulence: systematic and functional analyses of a pathogenicity island. *Proc. Natl. Acad. Sci. USA* 101:5597–5603.
- Fitzhenry, R. J., D. J. Pickard, E. L. Hartland, S. Reece, G. Dougan, A. D. Phillips, and G. Frankel. 2002. Intimin type influences the site of human intestinal mucosal colonisation by enterohaemorrhagic *Escherichia coli* O157:H7. *Gut* 51:180–185.
- Forrester, C., A. Darbouille-Michard, E. Wasch, C. Rich, E. Petat, F. Denis, and B. Joly. 1989. Adhesive properties of enteropathogenic *Escherichia coli* isolated from infants with acute diarrhea in Africa. *Eur. J. Clin. Microbiol. Infect. Dis.* 8:970–983.
- Frankel, G., A. D. Phillips, I. Rosenshine, G. Dougan, J. B. Kaper, and S. Knutton. 1998. Enteropathogenic and enterohemorrhagic *Escherichia coli*: more subversive elements. *Mol. Microbiol.* 29:911–921.
- Friedrich, A. W., M. Bielakowski, W. L. Zhang, M. Puls, T. Kuciukas, A. Ammon, and H. Karch. 2002. *Escherichia coli* harboring Shiga toxin 2 gene variants: frequency and association with clinical symptoms. *J. Infect. Dis.* 185:74–84.
- Gannon, V. P., C. Toering, S. A. Massi, and C. L. Gyles. 1990. Molecular cloning and nucleotide sequence of another variant of the *Escherichia coli* Shiga-like toxin II family. *J. Gen. Microbiol.* 136:1125–1135.
- Garmendia, J., G. Frankel, and V. Frepin. 2005. Enteropathogenic and enterohemorrhagic *Escherichia coli* infections: translocation, transfection, transformation. *Infect. Immun.* 73:2573–2585.
- Garmendia, J., A. D. Phillips, M. F. Carlier, Y. Chong, S. Schüller, O. Marches, S. Dahan, E. Oswald, R. K. Shaw, S. Knutton, and G. Frankel. 2004. TceE is an enterohemorrhagic *Escherichia coli* O157:H7 type III effector protein that couples Tir to the actin cytoskeleton. *Cell. Microbiol.* 6:1167–1183.
- Gruenheid, S., I. Sekirov, N. A. Thomas, W. Deng, P. O'Donnell, D. Goode, Y. Li, E. A. Frey, N. F. Brown, P. Metanikov, T. Pawson, K. Ashman, and B. H. Finlay. 2004. Identification and characterization of NleA, a non-LEE encoded type III translocated virulence factor of enterohaemorrhagic *Escherichia coli* O157:H7. *Mol. Microbiol.* 51:1233–1249.
- Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41:95–98.
- Hueck, C. J. 1998. Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol. Mol. Biol. Rev.* 62:137–153.
- Karnath, M. A., R. T. Steele, M. Petric, and C. Lim. 1983. Sporadic cases of haemolytic-uraemic syndrome associated with faecal cytotoxin and cytotoxin-producing *Escherichia coli* in stools. *Lancet* i:619–620.
- Kelly, M., E. Hart, R. Mundy, O. Marches, S. Wiles, L. Bades, S. Luck, M. Tashchek, G. Frankel, R. M. Robins-Brown, and E. L. Hartland. 2006. Essential role of the type III secretion system effector NleH in colonisation of mice by *Citrobacter rodentium*. *Infect. Immun.* 74:2326–2337.
- Knutton, S., D. R. Lloyd, and A. C. McNeill. 1987. Adhesion of enteropathogenic *Escherichia coli* to human intestinal enterocytes and cultured human intestinal mucosa. *Infect. Immun.* 55:69–77.
- Levine, M. M., E. J. Bergquist, D. R. Nalin, D. H. Waterman, R. B. Iorriick, C. R. Yomg, and S. Sotman. 1978. *Escherichia coli* strains that cause diarrhoea but do not produce heat labile or heat stable enterotoxins and are non-invasive. *Lancet* ii:1119–1122.
- Liu, Z., B. Kurasano, S. Tawakali, and Y. Takeda. 1993. Detection of various variant verotoxin genes in *Escherichia coli* by polymerase chain reaction. *Microbiol. Immunol.* 37:543–548.
- Marches, O., T. N. Ledger, M. Boury, M. Ohara, N. Tu, F. Goffaux, J. Mailu, I. Rosenshine, M. Sugai, J. De Rycke, and E. Oswald. 2003. Enteropathogenic and enterohaemorrhagic *Escherichia coli* deliver a novel effector called Cif, which blocks cell cycle G2/M transition. *Mol. Microbiol.* 50:1553–1567.
- Mariani-Kurkjian, P., E. Denamur, A. Milon, B. Picard, H. Caro, N. Lambert-Zechovsky, C. Lohr, P. Goulet, P. J. Sansonetti, and J. E. Hoon. 1993. Identification of a clone of *Escherichia coli* O103:H2 as a potential agent of hemolytic-uremic syndrome in France. *J. Clin. Microbiol.* 31:296–301.
- Mundy, R., C. Jenkins, J. Yu, H. Smith, and G. Frankel. 2004. Distribution of *espG* among clinical enterohaemorrhagic and enteropathogenic *Escherichia coli* isolates. *J. Med. Microbiol.* 53:145–149.
- Mundy, R., I. Petrovska, K. Smollett, N. Simpson, R. K. Wilson, J. Yu, X. Tu, I. Rosenshine, S. Clare, G. Dougan, and G. Frankel. 2004. Identification of a novel *Citrobacter rodentium* type III secreted protein, EspL, and roles of this and other secreted proteins in infection. *Infect. Immun.* 72:2289–2302.
- O'Brien, A. D., J. W. Newland, S. F. Miller, R. K. Holmes, H. W. Smith, and S. B. Formis. 1984. Shiga-like toxin-converting phages from *Escherichia coli* strains that cause hemorrhagic colitis or infantile diarrhea. *Science* 226:694–696.
- Oswald, E., H. Schmidt, S. Morabito, H. Karch, O. Marches, and A. Caprioli. 2000. Typing of intimin genes in human and animal enterohemorrhagic and enteropathogenic *Escherichia coli*: characterization of a new intimin variant. *Infect. Immun.* 68:604–71.
- Perna, N. T., G. Plunkett III, V. Portland, B. Mann, J. D. Glassner, D. J. Rose, G. P. Mayhew, P. S. Evans, J. Gray, H. G. Gargus, G. G. Pofai, J. Hackett, S. Klink, A. Boutin, Y. Shao, L. Miller, E. J. Grobbeck, N. W. Davis, A. Lim, E. T. Dimalanta, K. D. Potamoussi, A. Apodaca, T. S. Anantharaman, J. Lin, G. Yen, D. C. Schwartz, R. A. Welch, and F. R. Blattner. 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 409:529–533.
- Primrose, K. D. 1987. New and versatile cloning vectors with kanamycin-resistance marker. *Gene* 56:309–312.
- Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Schmidt, H., B. Plaszchke, S. Franko, H. Rüssmann, A. Schwarzkopf, J. Heesemann, and H. Karch. 1994. Differentiation in virulence patterns of *Escherichia coli* possessing *eae* genes. *Mol. Microbiol. Immunol.* 13:23–31.
- Schmidt, H., H. Rüssmann, and H. Karch. 1993. Virulence determinants in nonenterohemorrhagic *Escherichia coli* O157 strains that cause infantile diarrhea. *Infect. Immun.* 61:4894–4898.
- Schmidt, H., H. Rüssmann, A. Schwarzkopf, S. Aleksić, J. Heesemann, and H. Karch. 1994. Prevalence of attaching and effacing *Escherichia coli* in stool samples from patients and controls. *Zentralbl. Bakteriol.* 181:201–213.
- Schmidt, H., J. Scheef, S. Morabito, A. Caprioli, L. H. Weller, and H. Karch. 2000. A new Shiga toxin 2 variant (Stx2c) from *Escherichia coli* isolated from piglets. *Appl. Environ. Microbiol.* 66:1205–1212.
- Tobe, T., S. A. Beaton, H. Taniguchi, H. Abe, C. M. Bailey, A. Fritan, R. Yoniss, S. Matthews, O. Marches, G. Frankel, T. Hayashi, and M. J. Pollen. 2006. An extensive repertoire of type III secretion effectors in *Escherichia coli* O157 and the role of lambdoid phages in their dissemination. *Proc. Natl. Acad. Sci. USA* 103:14941–14946.

38. Vlisidou, I., O. Marches, F. Dava, R. Mundy, G. Frankol, and M. P. Stevens. 2006. Identification and characterization of EspK, a type III secreted effector protein of enterohemorrhagic *Escherichia coli* O157:H7. *FEMS Microbiol. Lett.* 263:32–40.
39. Voets, E., A. W. Paton, P. A. Manning, and J. C. Paton. 1998. Molecular analysis of Shiga toxinogenic *Escherichia coli* O111:H1 proteases which react with sera from patients with hemolytic-uremic syndrome. *Infect. Immun.* 66:1467–1472.
40. Zhang, W. L., B. Köbler, E. Oswald, L. Beutin, B. Karch, S. Morabito, A. Caprioli, S. Suerbaum, and H. Schmidt. 2002. Genetic diversity of antimin genes of attaching and effacing *Escherichia coli* strains. *J. Clin. Microbiol.* 40:4486–4492.

Appendix B

<Translation of JP2002-355074 A >

Pub. No.: p2002-355074 A
Publication Date: Dec. 10, 2002
5 Application No.: p2002-15959
Filing Date: Jan. 24, 2004
IPC: C12N 15/09 ZNA
A61K 31/7088
39/00
10 48/00
A61P 31/04
C07K 14/245
16/12
C12M 1/00
15 C12N 1/15
1/19
1/21
5/10
C12P 21/02
20 C12Q 1/68
G01N 33/15
33/50
33/53
33/566
25 37/00 102
Applicants: UNIVERSITY OF TSUKUBA; 1-1-1, Tennodai,
Tsukuba-shi, Ibaraki 3058577 (JP)
Inventors: HAYASHI Hideo et al. (JP)
Agent: TAKAGI Thiyosi et al.
30 Title: A nucleic-acid molecule and a polypeptide specific
to enterohemorrhagic E. coli O-157:H7 and a method
of using thereof
Priority Data: 2001-112010 Jan. 24, 2001 (JP)

35

CLAIMS

1. A nucleic-acid molecule specific to enterohemorrhagic pathogenic *E. coli* O-157:H7.

2. The nucleic-acid molecule of claim 1, which is a nucleic-acid molecule specific to enterohemorrhagic pathogenic *E. coli* O-157:H7 and has

(a) a nucleotide sequence selected from a group comprising the following SEQ IDs: SEQ ID NO:1, SEQ ID NO:132, SEQ ID NO:244, SEQ ID NO:337, SEQ ID NO:410, SEQ ID NO:484, SEQ ID NO:554, SEQ ID NO:630, SEQ ID NO:689, SEQ ID NO:755, SEQ ID NO:816, SEQ ID NO:876, SEQ ID NO:927, SEQ ID NO:978, SEQ ID NO:1013, SEQ ID NO:1029, SEQ ID NO:1055, SEQ ID NO:1060, SEQ ID NO:1093, SEQ ID NO:1128, SEQ ID NO:1157, SEQ ID NO:1191, SEQ ID NO:1212, SEQ ID NO:1240, SEQ ID NO:1258, SEQ ID NO:1274, SEQ ID NO:1288, SEQ ID NO:1302, SEQ ID NO:1309, SEQ ID NO:1321, SEQ ID NO:1329, SEQ ID NO:1338, SEQ ID NO:1348, SEQ ID NO:1359, SEQ ID NO:1366, SEQ ID NO:1374, SEQ ID NO:1380, SEQ ID NO:1386, SEQ ID NO:1394, SEQ ID NO:1401, SEQ ID NO:1408, SEQ ID NO:1411, SEQ ID NO:1418, SEQ ID NO:1426, SEQ ID NO:1436, SEQ ID NO:1443, SEQ ID NO:1450, SEQ ID NO:1457, SEQ ID NO:1460, SEQ ID NO:1467, SEQ ID NO:1471, SEQ ID NO:1473, SEQ ID NO:1478, SEQ ID NO:1487, SEQ ID NO:1489, SEQ ID NO:1494, SEQ ID NO:1499, SEQ ID NO:1501, SEQ ID NO:1506, SEQ ID NO:1508, SEQ ID NO:1510, SEQ ID NO:1511, SEQ ID NO:1516, SEQ ID NO:1520, SEQ ID NO:1526, SEQ ID NO:1532, SEQ ID NO:1537, SEQ ID NO:1540, SEQ ID NO:1545, SEQ ID NO:1547, SEQ ID NO:1549, SEQ ID NO:1551, SEQ ID NO:1553, SEQ ID NO:1555, SEQ ID NO:1558, SEQ ID NO:1563, SEQ ID NO:1566, SEQ ID NO:1569, SEQ ID NO:1571, SEQ ID NO:1576, SEQ ID NO:1580, SEQ ID NO:1584, SEQ ID NO:1587, SEQ ID NO:1591, SEQ ID NO:1594, SEQ ID NO:1596, SEQ ID NO:1599, SEQ ID NO:1601, SEQ ID NO:1603, SEQ ID NO:1604, SEQ ID NO:1605, SEQ ID NO:1607, SEQ ID

NO:1612, SEQ ID NO:1615, SEQ ID NO:1617, SEQ ID NO:1619,
70 SEQ ID NO:1622, SEQ ID NO:1624,SEQ ID NO:1626, SEQ ID
NO:1627, SEQ ID NO:1629, SEQ ID NO:1632, SEQID NO:1635,
SEQ ID NO:1636, SEQ ID NO:1637, SEQ ID NO:1639, SEQ
IDNO:1640, SEQ ID NO:1643, SEQ ID NO:1646, SEQ ID NO:
1649, SEQ ID NO:1652, SEQ ID NO:1655, SEQ ID NO:1658,
75 SEQ ID NO:1660, SEQ ID NO:1662, SEQ ID NO:1664, SEQ ID
NO:1666, SEQ ID NO:1668, SEQ ID NO:1669, SEQ ID NO:1670,
SEQ ID NO:1672, SEQ ID NO:1673, SEQ ID NO:1675, SEQ
IDNO:1677, SEQ ID NO:1680, SEQ ID NO:1682, SEQ ID NO:
1683, SEQ ID NO:1685, SEQ ID NO:1688, SEQ ID NO:1690,
80 SEQ ID NO:1691, SEQ ID NO:1694, SEQ ID NO:1696, SEQ ID
NO:1699, SEQ ID NO:1700, SEQ ID NO:1701,SEQ ID NO:1704,
SEQ ID NO:1705, SEQ ID NO:1706, SEQ ID NO:1707, SEQID
NO:1708, SEQ ID NO:1709, SEQ ID NO:1710, SEQ ID NO:1711,
SEQ ID NO:1712, SEQ ID NO:1713, SEQ ID NO:1715, SEQ ID
85 NO:1716, SEQ ID NO:1717, SEQ ID NO:1718,, SEQ ID NO:
1719, SEQ ID NO:1720, SEQ ID NO:1721, SEQ ID NO:1722,
SEQ ID NO:1723, SEQ ID NO:1724, SEQ ID NO:1725, SEQ ID
NO:1726, SEQ ID NO:1727, SEQ ID NO:1728, SEQ ID NO:1729,
SEQ IDNO:1730, SEQ ID NO:1731, SEQ ID NO:1732, SEQ ID
90 NO:1733, SEQ ID NO:1734, SEQ ID NO:1735, SEQ ID NO:1736,
SEQ ID NO:1737, SEQ ID NO:1738, SEQ ID NO:1739, SEQ ID
NO:1740, SEQ ID NO:1741, SEQ ID NO:1742,SEQ ID NO:1743,
SEQ ID NO:1744, SEQ ID NO:1745, SEQ ID NO:1746, SEQID
NO:1747, SEQ ID NO:1748, SEQ ID NO:1749, SEQ ID NO:1750.
95 SEQ ID NO:1751, SEQ ID NO:1752, SEQ ID NO:1753, SEQ ID
NO:1754, SEQ ID NO:1755, SEQ ID NO:1756, SEQ ID NO:1757,
SEQ ID NO:1758, SEQ ID NO:1759,SEQ ID NO:1760, SEQ ID
NO:1761, SEQ ID NO:1762, SEQ ID NO:1763, SEQID NO:1764,
SEQ ID NO:1765, SEQ ID NO:1766, SEQ ID NO:1767, SEQ
100 IDNO:1768, SEQ ID NO:1769, SEQ ID NO:1770, SEQ ID NO:
1771, SEQ ID NO:1772, SEQ ID NO:1773, SEQ ID NO:1774,
SEQ ID NO:1775, SEQ ID NO:1776, SEQ ID NO:1777, SEQ ID

NO:1778, SEQ ID NO:1779, SEQ ID NO:1780, SEQ ID NO:1781,
SEQ ID NO:1782, SEQ ID NO:1783, SEQ ID NO:1784, SEQ
105 IDNO:1785, SEQ ID NO:1786, SEQ ID NO:1787, SEQ ID NO:
1788, SEQ ID NO:1789, SEQ ID NO:1790, SEQ ID NO:1791,
SEQ ID NO:1792, SEQ ID NO:1793, SEQ ID NO:1794, SEQ ID
NO:1795, SEQ ID NO:1796, SEQ ID NO:1797,SEQ ID NO:1798,
SEQ ID NO:1799, SEQ ID NO:1800, SEQ ID NO:1801, SEQID
110 NO:1802, SEQ ID NO:1803, SEQ ID NO:1804, SEQ ID NO:1805,
SEQ ID NO:1806, SEQ ID NO:1807, SEQ ID NO:1808, SEQ ID
NO:1809, SEQ ID NO:1810, SEQ ID NO:1811, SEQ ID NO:1812,
SEQ ID NO:1813, SEQ ID NO:1814,SEQ ID NO:1815, SEQ ID
NO:1816, SEQ ID NO:1817, SEQ ID NO:1818, SEQID NO:1819,
115 SEQ ID NO:1820, SEQ ID NO:1821, SEQ ID NO:1822, SEQ
IDNO:1823, SEQ ID NO:1824, SEQ ID NO:1825, SEQ ID NO:
1826, SEQ ID NO:1827, SEQ ID NO:1828, SEQ ID NO:1829,
SEQ ID NO:1830, SEQ ID NO:1831, SEQ ID NO:1832, SEQ ID
NO:1833, SEQ ID NO:1834, SEQ ID NO:1835, SEQ ID NO:1836,
120 SEQ ID NO:1837, SEQ ID NO:1838, SEQ ID NO:1839, SEQ
IDNO:1840, SEQ ID NO:1841, SEQ ID NO:1842, SEQ ID NO:
1843, SEQ ID NO:1844, SEQ ID NO:1845, SEQ ID NO:1846,
SEQ ID NO:1847, SEQ ID NO:1848, SEQ ID NO:1849, SEQ ID
NO:1850, SEQ ID NO:1851, SEQ ID NO:1852,SEQ ID NO:1853,
125 SEQ ID NO:1854, SEQ ID NO:1855, SEQ ID NO:1856, SEQID
NO:1857, SEQ ID NO:1858, SEQ ID NO:1859, SEQ ID NO:1860,
SEQ ID NO:1861, SEQ ID NO:1862, SEQ ID NO:1863, SEQ ID
NO:1864, SEQ ID NO:1865, and SEQ ID NO:1866

(b) a moiety in the nucleotide sequences set forth in (a):

130 (c) a complementary nucleotide sequence to the
nucleotide sequences set forth in (a) or (b); or

(d) a nucleotide sequence hybridizing to the nucleotide
sequences set forth in (a), (b) or (c) under a stringent condition.

3. The nucleic-acid molecule of claim 1, which is a
135 nucleic-acid molecule encoding a polypeptide specific to
enterohemorrhagic pathogenic-E. coli O-157:H7 and encodes

(a) an amino acid sequence selected from a group comprising the following SEQ IDs or a moiety thereof: SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ IDNO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25,SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ IDNO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49,SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73,SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ IDNO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97,SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117, SEQ ID NO:118, SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID

NO:128, SEQ ID NO:129, SEQ ID NO:130, SEQ ID NO:131,
 SEQ ID NO:133, SEQ ID NO:134, SEQ ID NO:135, SEQ ID NO:
 136, SEQ ID NO:137, SEQ ID NO:138, SEQ ID NO:139, SEQ ID
 175 NO:140, SEQ ID NO:141, SEQ ID NO:142, SEQ ID NO:143,
 SEQ ID NO:144, SEQ ID NO:145, SEQ ID NO:146, SEQ ID NO:
 147, SEQ ID NO:148, SEQ ID NO:149, SEQ ID NO:150, SEQ ID
 NO:151, SEQ ID NO:152, SEQ ID NO:153, SEQ ID NO:154,
 SEQ ID NO:155, SEQ ID NO:156, SEQ ID NO:157, SEQ ID NO:
 158, SEQ ID NO:159, SEQ ID NO:160, SEQ ID NO:161, SEQ ID
 180 NO:162, SEQ ID NO:163, SEQ ID NO:164, SEQ ID NO:165,
 SEQ ID NO:166, SEQ ID NO:167, SEQ ID NO:168, SEQ ID NO:
 169, SEQ ID NO:170, SEQ ID NO:171, SEQ ID NO:172, SEQ ID
 NO:173, SEQ ID NO:174, SEQ ID NO:175, SEQ ID NO:176,
 SEQ ID NO:177, SEQ ID NO:178, SEQ ID NO:179, SEQ ID NO:
 185 180, SEQ ID NO:181, SEQ ID NO:182, SEQ ID NO:183, SEQ ID
 NO:184, SEQ ID NO:185, SEQ ID NO:186, SEQ ID NO:187,
 SEQ ID NO:188, SEQ ID NO:189, SEQ ID NO:190, SEQ ID NO:
 191, SEQ ID NO:192, SEQ ID NO:193, SEQ ID NO:194, SEQ ID
 NO:195, SEQ ID NO:196, SEQ ID NO:197, SEQ ID NO:198,
 190 SEQ ID NO:199, SEQ ID NO:200, SEQ ID NO:201, SEQ ID NO:
 202, SEQ ID NO:203, SEQ ID NO:204, SEQ ID NO:205, SEQ ID
 NO:206, SEQ ID NO:207, SEQ ID NO:208, SEQ ID NO:209,
 SEQ ID NO:210, SEQ ID NO:211, SEQ ID NO:212, SEQ ID NO:
 213, SEQ ID NO:214, SEQ ID NO:215, SEQ ID NO:216, SEQ ID
 195 NO:217, SEQ ID NO:218, SEQ ID NO:219, SEQ ID NO:220,
 SEQ ID NO:221, SEQ ID NO:222, SEQ ID NO:223, SEQ ID NO:
 224, SEQ ID NO:225, SEQ ID NO:226, SEQ ID NO:227, SEQ ID
 NO:228, SEQ ID NO:229, SEQ ID NO:230, SEQ ID NO:231,
 SEQ ID NO:232, SEQ ID NO:233, SEQ ID NO:234, SEQ ID NO:
 200 235, SEQ ID NO:236, SEQ ID NO:237, SEQ ID NO:238, SEQ ID
 NO:239, SEQ ID NO:240, SEQ ID NO:241, SEQ ID NO:242,
 SEQ ID NO:243, SEQ ID NO:245, SEQ ID NO:246, SEQ ID NO:
 247, SEQ ID NO:248, SEQ ID NO:249, SEQ ID NO:250, SEQ ID
 NO:251, SEQ ID NO:252, SEQ ID NO:253, SEQ ID NO:254,

205 SEQ ID NO:255, SEQ ID NO:256, SEQ ID NO:257, SEQ ID NO:
258, SEQ ID NO:259, SEQ ID NO:260, SEQ ID NO:261, SEQ ID
NO:262, SEQ ID NO:263, SEQ ID NO:264, SEQ ID NO:265,
SEQ ID NO:266, SEQ ID NO:267, SEQ ID NO:268, SEQ ID NO:
210 NO:269, SEQ ID NO:270, SEQ ID NO:271, SEQ ID NO:272, SEQ ID
NO:273, SEQ ID NO:274, SEQ ID NO:275, SEQ ID NO:276,
SEQ ID NO:277, SEQ ID NO:278, SEQ ID NO:279, SEQ ID NO:
280, SEQ ID NO:281, SEQ ID NO:282, SEQ ID NO:283, SEQ ID
NO:284, SEQ ID NO:285, SEQ ID NO:286, SEQ ID NO:287,
SEQ ID NO:288, SEQ ID NO:289, SEQ ID NO:290, SEQ ID NO:
215 291, SEQ ID NO:292, SEQ ID NO:293, SEQ ID NO:294, SEQ ID
NO:295, SEQ ID NO:296, SEQ ID NO:297, SEQ ID NO:298,
SEQ ID NO:299, SEQ ID NO:300, SEQ ID NO:301, SEQ ID NO:
302, SEQ ID NO:303, SEQ ID NO:304, SEQ ID NO:305, SEQ ID
NO:306, SEQ ID NO:307, SEQ ID NO:308, SEQ ID NO:309,
220 SEQ ID NO:310, SEQ ID NO:311, SEQ ID NO:312, SEQ ID NO:
313, SEQ ID NO:314, SEQ ID NO:315, SEQ ID NO:316, SEQ ID
NO:317, SEQ ID NO:318, SEQ ID NO:319, SEQ ID NO:320,
SEQ ID NO:321, SEQ ID NO:322, SEQ ID NO:323, SEQ ID NO:
324, SEQ ID NO:325, SEQ ID NO:326, SEQ ID NO:327, SEQ ID
225 NO:328, SEQ ID NO:329, SEQ ID NO:330, SEQ ID NO:331,
SEQ ID NO:332, SEQ ID NO:333, SEQ ID NO:334, SEQ ID NO:
335, SEQ ID NO:336, SEQ ID NO:338, SEQ ID NO:339, SEQ ID
NO:340, SEQ ID NO:341, SEQ ID NO:342, SEQ ID NO:343,
SEQ ID NO:344, SEQ ID NO:345, SEQ ID NO:346, SEQ ID NO:
230 347, SEQ ID NO:348, SEQ ID NO:349, SEQ ID NO:350, SEQ ID
NO:351, SEQ ID NO:352, SEQ ID NO:353, SEQ ID NO:354,
SEQ ID NO:355, SEQ ID NO:356, SEQ ID NO:357, SEQ ID NO:
358, SEQ ID NO:359, SEQ ID NO:360, SEQ ID NO:361, SEQ ID
NO:362, SEQ ID NO:363, SEQ ID NO:364, SEQ ID NO:365,
235 SEQ ID NO:366, SEQ ID NO:367, SEQ ID NO:368, SEQ ID NO:
369, SEQ ID NO:370, SEQ ID NO:371, SEQ ID NO:372, SEQ ID
NO:373, SEQ ID NO:374, SEQ ID NO:375, SEQ ID NO:376,
SEQ ID NO:377, SEQ ID NO:378, SEQ ID NO:379, SEQ ID NO:

380, SEQ ID NO:381, SEQ ID NO:382, SEQ ID NO:383, SEQ ID
240 NO:384, SEQ ID NO:385, SEQ ID NO:386, SEQ ID NO:387,
SEQ ID NO:388, SEQ ID NO:389, SEQ ID NO:390, SEQ ID NO:
391, SEQ ID NO:392, SEQ ID NO:393, SEQ ID NO:394, SEQ ID
NO:395, SEQ ID NO:396, SEQ ID NO:397, SEQ ID NO:398,
SEQ ID NO:399, SEQ ID NO:400, SEQ ID NO:401, SEQ ID NO:
245 402, SEQ ID NO:403, SEQ ID NO:404, SEQ ID NO:405, SEQ ID
NO:406, SEQ ID NO:407, SEQ ID NO:408, SEQ ID NO:409,
SEQ ID NO:411, SEQ ID NO:412, SEQ ID NO:413, SEQ ID NO:
414, SEQ ID NO:415, SEQ ID NO:416, SEQ ID NO:417, SEQ ID
NO:418, SEQ ID NO:419, SEQ ID NO:420, SEQ ID NO:421,
250 SEQ ID NO:422, SEQ ID NO:423, SEQ ID NO:424, SEQ ID NO:
425, SEQ ID NO:426, SEQ ID NO:427, SEQ ID NO:428, SEQ ID
NO:429, SEQ ID NO:430, SEQ ID NO:431, SEQ ID NO:432,
SEQ ID NO:433, SEQ ID NO:434, SEQ ID NO:435, SEQ ID NO:
436, SEQ ID NO:437, SEQ ID NO:438, SEQ ID NO:439, SEQ ID
255 NO:440, SEQ ID NO:441, SEQ ID NO:442, SEQ ID NO:443,
SEQ ID NO:444, SEQ ID NO:445, SEQ ID NO:446, SEQ ID NO:
447, SEQ ID NO:448, SEQ ID NO:449, SEQ ID NO:450, SEQ ID
NO:451, SEQ ID NO:452, SEQ ID NO:453, SEQ ID NO:454,
SEQ ID NO:455, SEQ ID NO:456, SEQ ID NO:457, SEQ ID NO:
260 458, SEQ ID NO:459, SEQ ID NO:460, SEQ ID NO:461, SEQ ID
NO:462, SEQ ID NO:463, SEQ ID NO:464, SEQ ID NO:465,
SEQ ID NO:466, SEQ ID NO:467, SEQ ID NO:468, SEQ ID NO:
469, SEQ ID NO:470, SEQ ID NO:471, SEQ ID NO:472, SEQ ID
NO:473, SEQ ID NO:474, SEQ ID NO:475, SEQ ID NO:476,
265 SEQ ID NO:477, SEQ ID NO:478, SEQ ID NO:479, SEQ ID NO:
480, SEQ ID NO:481, SEQ ID NO:482, SEQ ID NO:483, SEQ ID
NO:485, SEQ ID NO:486, SEQ ID NO:487, SEQ ID NO:488,
SEQ ID NO:489, SEQ ID NO:490, SEQ ID NO:491, SEQ ID NO:
492, SEQ ID NO:493, SEQ ID NO:494, SEQ ID NO:495, SEQ ID
270 NO:496, SEQ ID NO:497, SEQ ID NO:498, SEQ ID NO:499,
SEQ ID NO:500, SEQ ID NO:501, SEQ ID NO:502, SEQ ID NO:
503, SEQ ID NO:504, SEQ ID NO:505, SEQ ID NO:506, SEQ ID

NO:507, SEQ ID NO:508, SEQ ID NO:509, SEQ ID NO:510,
SEQ ID NO:511, SEQ ID NO:512, SEQ ID NO:513, SEQ ID NO:
275 514, SEQ ID NO:515, SEQ ID NO:516, SEQ ID NO:517, SEQ ID
NO:518, SEQ ID NO:519, SEQ ID NO:520, SEQ ID NO:521,
SEQ ID NO:522, SEQ ID NO:523, SEQ ID NO:524, SEQ ID NO:
525, SEQ ID NO:526, SEQ ID NO:527, SEQ ID NO:528, SEQ ID
NO:529, SEQ ID NO:530, SEQ ID NO:531, SEQ ID NO:532,
280 SEQ ID NO:533, SEQ ID NO:534, SEQ ID NO:535, SEQ ID
NO:536, SEQ ID NO:537, SEQ ID NO:538, SEQ ID NO:539,
SEQ ID NO:540, SEQ ID NO:541, SEQ ID NO:542, SEQ ID NO:
543, SEQ ID NO:544, SEQ ID NO:545, SEQ ID NO:546, SEQ ID
NO:547, SEQ ID NO:548, SEQ ID NO:549, SEQ ID NO:550,
285 SEQ ID NO:551, SEQ ID NO:552, SEQ ID NO:553, SEQ ID NO:
555, SEQ ID NO:556, SEQ ID NO:557, SEQ ID NO:558, SEQ ID
NO:559, SEQ ID NO:560, SEQ ID NO:561, SEQ ID NO:562,
SEQ ID NO:563, SEQ ID NO:564, SEQ ID NO:565, SEQ ID NO:
566, SEQ ID NO:567, SEQ ID NO:568, SEQ ID NO:569, SEQ ID
290 NO:570, SEQ ID NO:571, SEQ ID NO:572, SEQ ID NO:573,
SEQ ID NO:574, SEQ ID NO:575, SEQ ID NO:576, SEQ ID NO:
577, SEQ ID NO:578, SEQ ID NO:579, SEQ ID NO:580, SEQ ID
NO:581, SEQ ID NO:582, SEQ ID NO:583, SEQ ID NO:584,
SEQ ID NO:585, SEQ ID NO:586, SEQ ID NO:587, SEQ ID NO:
295 588, SEQ ID NO:589, SEQ ID NO:590, SEQ ID NO:591, SEQ ID
NO:592, SEQ ID NO:593, SEQ ID NO:594, SEQ ID NO:595,
SEQ ID NO:596, SEQ ID NO:597, SEQ ID NO:598, SEQ ID NO:
599, SEQ ID NO:600, SEQ ID NO:601, SEQ ID NO:602, SEQ ID
NO:603, SEQ ID NO:604, SEQ ID NO:605, SEQ ID NO:606,
300 SEQ ID NO:607, SEQ ID NO:608, SEQ ID NO:609, SEQ ID NO:
610, SEQ ID NO:611, SEQ ID NO:612, SEQ ID NO:613, SEQ ID
NO:614, SEQ ID NO:615, SEQ ID NO:616, SEQ ID NO:617,
SEQ ID NO:618, SEQ ID NO:619, SEQ ID NO:620, SEQ ID NO:
621, SEQ ID NO:622, SEQ ID NO:623, SEQ ID NO:624, SEQ ID
305 NO:625, SEQ ID NO:626, SEQ ID NO:627, SEQ ID NO:628,
SEQ ID NO:629, SEQ ID NO:631, SEQ ID NO:632, SEQ ID NO:

633, SEQ ID NO:634, SEQ ID NO:635, SEQ ID NO:636, SEQ ID
NO:637, SEQ ID NO:638, SEQ ID NO:639, SEQ ID NO:640,
SEQ ID NO:641, SEQ ID NO:642, SEQ ID NO:643, SEQ ID NO:
310 644, SEQ ID NO:645, SEQ ID NO:646, SEQ ID NO:647, SEQ ID
NO:648, SEQ ID NO:649, SEQ ID NO:650, SEQ ID NO:651,
SEQ ID NO:652, SEQ ID NO:653, SEQ ID NO:654, SEQ ID NO:
655, SEQ ID NO:656, SEQ ID NO:657, SEQ ID NO:658, SEQ ID
NO:659, SEQ ID NO:660, SEQ ID NO:661, SEQ ID NO:662,
315 SEQ ID NO:663, SEQ ID NO:664, SEQ ID NO:665, SEQ ID NO:
666, SEQ ID NO:667, SEQ ID NO:668, SEQ ID NO:669, SEQ ID
NO:670, SEQ ID NO:671, SEQ ID NO:672, SEQ ID NO:673,
SEQ ID NO:674, SEQ ID NO:675, SEQ ID NO:676, SEQ ID NO:
677, SEQ ID NO:678, SEQ ID NO:679, SEQ ID NO:680, SEQ ID
320 NO:681, SEQ ID NO:682, SEQ ID NO:683, SEQ ID NO:684,
SEQ ID NO:685, SEQ ID NO:686, SEQ ID NO:687, SEQ ID NO:
688, SEQ ID NO:690, SEQ ID NO:691, SEQ ID NO:692, SEQ ID
NO:693, SEQ ID NO:694, SEQ ID NO:695, SEQ ID NO:696,
SEQ ID NO:697, SEQ ID NO:698, SEQ ID NO:699, SEQ ID NO:
325 700, SEQ ID NO:701, SEQ ID NO:702, SEQ ID NO:703, SEQ ID
NO:704, SEQ ID NO:705, SEQ ID NO:706, SEQ ID NO:707,
SEQ ID NO:708, SEQ ID NO:709, SEQ ID NO:710, SEQ ID NO:
711, SEQ ID NO:712, SEQ ID NO:713, SEQ ID NO:714, SEQ ID
NO:715, SEQ ID NO:716, SEQ ID NO:717, SEQ ID NO:718,
330 SEQ ID NO:719, SEQ ID NO:720, SEQ ID NO:721, SEQ ID NO:
722, SEQ ID NO:723, SEQ ID NO:724, SEQ ID NO:725, SEQ ID
NO:726, SEQ ID NO:727, SEQ ID NO:728, SEQ ID NO:729,
SEQ ID NO:730, SEQ ID NO:731, SEQ ID NO:732, SEQ ID NO:
733, SEQ ID NO:734, SEQ ID NO:735, SEQ ID NO:736, SEQ ID
335 NO:737, SEQ ID NO:738, SEQ ID NO:739, SEQ ID NO:740,
SEQ ID NO:741, SEQ ID NO:742, SEQ ID NO:743, SEQ ID NO:
744, SEQ ID NO:745, SEQ ID NO:746, SEQ ID NO:747, SEQ ID
NO:748, SEQ ID NO:749, SEQ ID NO:750, SEQ ID NO:751,
SEQ ID NO:752, SEQ ID NO:753, SEQ ID NO:754, SEQ ID NO:
340 756, SEQ ID NO:757, SEQ ID NO:758, SEQ ID NO:759, SEQ ID

NO:760, SEQ ID NO:761, SEQ ID NO:762, SEQ ID NO:763,
SEQ ID NO:764, SEQ ID NO:765, SEQ ID NO:766, SEQ ID NO:
767, SEQ ID NO:768, SEQ ID NO:769, SEQ ID NO:770, SEQ ID
NO:771, SEQ ID NO:772, SEQ ID NO:773, SEQ ID NO:774,
345 SEQ ID NO:775, SEQ ID NO:776, SEQ ID NO:777, SEQ ID NO:
778, SEQ ID NO:779, SEQ ID NO:780, SEQ ID NO:781, SEQ ID
NO:782, SEQ ID NO:783, SEQ ID NO:784, SEQ ID NO:785,
SEQ ID NO:786, SEQ ID NO:787, SEQ ID NO:788, SEQ ID NO:
789, SEQ ID NO:790, SEQ ID NO:791, SEQ ID NO:792, SEQ ID
350 NO:793, SEQ ID NO:794, SEQ ID NO:795, SEQ ID NO:796,
SEQ ID NO:797, SEQ ID NO:798, SEQ ID NO:799, SEQ ID N
O:800, SEQ ID NO:801, SEQ ID NO:802, SEQ ID NO:803, SEQ
ID NO:804, SEQ ID NO:805, SEQ ID NO:806, SEQ ID NO:807,
SEQ ID NO:808, SEQ ID NO:809, SEQ ID NO:810, SEQ ID NO:
355 811, SEQ ID NO:812, SEQ ID NO:813, SEQ ID NO:814, SEQ ID
NO:815, SEQ ID NO:817, SEQ ID NO:818, SEQ ID NO:819,
SEQ ID NO:820, SEQ ID NO:821, SEQ ID NO:822, SEQ ID NO:
823, SEQ ID NO:824, SEQ ID NO:825, SEQ ID NO:826, SEQ ID
NO:827, SEQ ID NO:828, SEQ ID NO:829, SEQ ID NO:830,
360 SEQ ID NO:831, SEQ ID NO:832, SEQ ID NO:833, SEQ ID NO:
834, SEQ ID NO:835, SEQ ID NO:836, SEQ ID NO:837, SEQ ID
NO:838, SEQ ID NO:839, SEQ ID NO:840, SEQ ID NO:841,
SEQ ID NO:842, SEQ ID NO:843, SEQ ID NO:844, SEQ ID NO:
845, SEQ ID NO:846, SEQ ID NO:847, SEQ ID NO:848, SEQ ID
365 NO:849, SEQ ID NO:850, SEQ ID NO:851, SEQ ID NO:852,
SEQ ID NO:853, SEQ ID NO:854, SEQ ID NO:855, SEQ ID NO:
856, SEQ ID NO:857, SEQ ID NO:858, SEQ ID NO:859, SEQ ID
NO:860, SEQ ID NO:861, SEQ ID NO:862, SEQ ID NO:863,
SEQ ID NO:864, SEQ ID NO:865, SEQ ID NO:866, SEQ ID NO:
370 867, SEQ ID NO:868, SEQ ID NO:869, SEQ ID NO:870, SEQ ID
NO:871, SEQ ID NO:872, SEQ ID NO:873, SEQ ID NO:874,
SEQ ID NO:875, SEQ ID NO:877, SEQ ID NO:878, SEQ ID NO:
879, SEQ ID NO:880, SEQ ID NO:881, SEQ ID NO:882, SEQ ID
NO:883, SEQ ID NO:884, SEQ ID NO:885, SEQ ID NO:886,

375 SEQ ID NO:887, SEQ ID NO:888, SEQ ID NO:889, SEQ ID NO:
 890, SEQ ID NO:891, SEQ ID NO:892, SEQ ID NO:893, SEQ ID
 NO:894, SEQ ID NO:895, SEQ ID NO:896, SEQ ID NO:897,
 SEQ ID NO:898, SEQ ID NO:899, SEQ ID NO:900, SEQ ID NO:
 901, SEQ ID NO:902, SEQ ID NO:903, SEQ ID NO:904, SEQ ID
 380 NO:905, SEQ ID NO:906, SEQ ID NO:907, SEQ ID NO:908,
 SEQ ID NO:909, SEQ ID NO:910, SEQ ID NO:911, SEQ ID NO:
 912, SEQ ID NO:913, SEQ ID NO:914, SEQ ID NO:915, SEQ ID
 NO:916, SEQ ID NO:917, SEQ ID NO:918, SEQ ID NO:919,
 SEQ ID NO:920, SEQ ID NO:921, SEQ ID NO:922, SEQ ID NO:
 385 923, SEQ ID NO:924, SEQ ID NO:925, SEQ ID NO:926, SEQ ID
 NO:928, SEQ ID NO:929, SEQ ID NO:930, SEQ ID NO:931,
 SEQ ID NO:932, SEQ ID NO:933, SEQ ID NO:934, SEQ ID NO:
 935, SEQ ID NO:936, SEQ ID NO:937, SEQ ID NO:938, SEQ ID
 NO:939, SEQ ID NO:940, SEQ ID NO:941, SEQ ID NO:942,
 390 SEQ ID NO:943, SEQ ID NO:944, SEQ ID NO:945, SEQ ID NO:
 946, SEQ ID NO:947, SEQ ID NO:948, SEQ ID NO:949, SEQ ID
 NO:950, SEQ ID NO:951, SEQ ID NO:952, SEQ ID NO:953,
 SEQ ID NO:954, SEQ ID NO:955, SEQ ID NO:956, SEQ ID NO:
 957, SEQ ID NO:958, SEQ ID NO:959, SEQ ID NO:960, SEQ ID
 395 NO:961, SEQ ID NO:962, SEQ ID NO:963, SEQ ID NO:964,
 SEQ ID NO:965, SEQ ID NO:966, SEQ ID NO:967, SEQ ID NO:
 968, SEQ ID NO:969, SEQ ID NO:970, SEQ ID NO:971, SEQ ID
 NO:972, SEQ ID NO:973, SEQ ID NO:974, SEQ ID NO:975,
 SEQ ID NO:976, SEQ ID NO:977, SEQ ID NO:979, SEQ ID NO:
 400 980, SEQ ID NO:981, SEQ ID NO:982, SEQ ID NO:983, SEQ ID
 NO:984, SEQ ID NO:985, SEQ ID NO:986, SEQ ID NO:987,
 SEQ ID NO:988, SEQ ID NO:989, SEQ ID NO:990, SEQ ID NO:
 991, SEQ ID NO:992, SEQ ID NO:993, SEQ ID NO:994, SEQ ID
 NO:995, SEQ ID NO:996, SEQ ID NO:997, SEQ ID NO:998,
 405 SEQ ID NO:999, SEQ ID NO:1000, SEQ ID NO:1001, SEQ ID
 NO:1002, SEQ ID NO:1003, SEQ ID NO:1004, SEQ ID NO:1005,
 SEQ ID NO:1006, SEQ ID NO:1007, SEQ ID NO:1008, SEQ ID
 NO:1009, SEQ ID NO:1010, SEQ ID NO:1011, SEQ ID NO:1012,

SEQ ID NO:1014, SEQ ID NO:1015, SEQ ID NO:1016, SEQ ID
410 NO:1017, SEQ ID NO:1018, SEQ ID NO:1019, SEQ ID NO:1020,
SEQ ID NO:1021, SEQ ID NO:1022, SEQ ID NO:1023, SEQ ID
NO:1024, SEQ ID NO:1025, SEQ ID NO:1026, SEQ ID NO:1027,
SEQ IDNO:1028, SEQ ID NO:1030, SEQ ID NO:1031, SEQ ID
NO:1032, SEQ ID NO:1033, SEQ ID NO:1034, SEQ ID NO:1035,
415 SEQ ID NO:1036, SEQ ID NO:1037, SEQ ID NO:1038, SEQ ID
NO:1039, SEQ ID NO:1040, SEQ ID NO:1041,SEQ ID NO:1042,
SEQ ID NO:1043, SEQ ID NO:1044, SEQ ID NO:1045, SEQID
NO:1046, SEQ ID NO:1047, SEQ ID NO:1048, SEQ ID NO:1049,
SEQ ID NO:1050, SEQ ID NO:1051, SEQ ID NO:1052, SEQ ID
420 NO:1053, SEQ ID NO:1054, SEQ ID NO:1056, SEQ ID NO:1057,
SEQ ID NO:1058, SEQ ID NO:1059,SEQ ID NO:1061, SEQ ID
NO:1062, SEQ ID NO:1063, SEQ ID NO:1064, SEQID NO:1065,
SEQ ID NO:1066, SEQ ID NO:1067, SEQ ID NO:1068, SEQ
IDNO:1069, SEQ ID NO:1070, SEQ ID NO:1071, SEQ ID NO:
425 1072, SEQ ID NO:1073, SEQ ID NO:1074, SEQ ID NO:1075,
SEQ ID NO:1076, SEQ ID NO:1077, SEQ ID NO:1078, SEQ ID
NO:1079, SEQ ID NO:1080, SEQ ID NO:1081, SEQ ID NO:1082,
SEQ ID NO:1083, SEQ ID NO:1084, SEQ ID NO:1085, SEQ
IDNO:1086, SEQ ID NO:1087, SEQ ID NO:1088, SEQ ID NO:
430 1089, SEQ ID NO:1090, SEQ ID NO:1091, SEQ ID NO:1092,
SEQ ID NO:1094, SEQ ID NO:1095, SEQ ID NO:1096, SEQ ID
NO:1097, SEQ ID NO:1098, SEQ ID NO:1099,SEQ ID NO:1100,
SEQ ID NO:1101, SEQ ID NO:1102, SEQ ID NO:1103, SEQID
NO:1104, SEQ ID NO:1105, SEQ ID NO:1106, SEQ ID NO:1107,
435 SEQ ID NO:1108, SEQ ID NO:1109, SEQ ID NO:1110, SEQ ID
NO:1111, SEQ ID NO:1112, SEQ ID NO:1113, SEQ ID NO:1114,
SEQ ID NO:1115, SEQ ID NO:1116,SEQ ID NO:1117, SEQ ID
NO:1118, SEQ ID NO:1119, SEQ ID NO:1120, SEQID NO:1121,
SEQ ID NO:1122, SEQ ID NO:1123, SEQ ID NO:1124, SEQ
440 IDNO:1125, SEQ ID NO:1126, SEQ ID NO:1127, SEQ ID NO:
1129, SEQ ID NO:1130, SEQ ID NO:1131, SEQ ID NO:1132,
SEQ ID NO:1133, SEQ ID NO:1134, SEQ ID NO:1135, SEQ ID

NO:1136, SEQ ID NO:1137, SEQ ID NO:1138, SEQ ID NO:1139,
 SEQ ID NO:1140, SEQ ID NO:1141, SEQ ID NO:1142, SEQ
 445 IDNO:1143, SEQ ID NO:1144, SEQ ID NO:1145, SEQ ID NO:
 1146, SEQ ID NO:1147, SEQ ID NO:1148, SEQ ID NO:1149,
 SEQ ID NO:1150, SEQ ID NO:1151, SEQ ID NO:1152, SEQ ID
 NO:1153, SEQ ID NO:1154, SEQ ID NO:1155,SEQ ID NO:1156,
 SEQ ID NO:1158, SEQ ID NO:1159, SEQ ID NO:1160, SEQID
 450 NO:1161, SEQ ID NO:1162, SEQ ID NO:1163, SEQ ID NO:1164,
 SEQ ID NO:1165, SEQ ID NO:1166, SEQ ID NO:1167, SEQ ID
 NO:1168, SEQ ID NO:1169, SEQ ID NO:1170, SEQ ID NO:1171,
 SEQ ID NO:1172, SEQ ID NO:1173,SEQ ID NO:1174, SEQ ID
 NO:1175, SEQ ID NO:1176, SEQ ID NO:1177, SEQID NO:1178,
 455 SEQ ID NO:1179, SEQ ID NO:1180, SEQ ID NO:1181, SEQ
 IDNO:1182, SEQ ID NO:1183, SEQ ID NO:1184, SEQ ID NO:
 1185, SEQ ID NO:1186, SEQ ID NO:1187, SEQ ID NO:1188,
 SEQ ID NO:1189, SEQ ID NO:1190, SEQ ID NO:1192, SEQ ID
 NO:1193, SEQ ID NO:1194, SEQ ID NO:1195, SEQ ID NO:1196,
 460 SEQ ID NO:1197, SEQ ID NO:1198, SEQ ID NO:1199, SEQ
 IDNO:1200, SEQ ID NO:1201, SEQ ID NO:1202, SEQ ID NO:
 1203, SEQ ID NO:1204, SEQ ID NO:1205, SEQ ID NO:1206,
 SEQ ID NO:1207, SEQ ID NO:1208, SEQ ID NO:1209, SEQ ID
 NO:1210, SEQ ID NO:1211, SEQ ID NO:1213,SEQ ID NO:1214,
 465 SEQ ID NO:1215, SEQ ID NO:1216, SEQ ID NO:1217, SEQID
 NO:1218, SEQ ID NO:1219, SEQ ID NO:1220, SEQ ID NO:1221,
 SEQ ID NO:1222, SEQ ID NO:1223, SEQ ID NO:1224, SEQ ID
 NO:1225, SEQ ID NO:1226, SEQ ID NO:1227, SEQ ID NO:1228,
 SEQ ID NO:1229, SEQ ID NO:1230,SEQ ID NO:1231, SEQ ID
 470 NO:1232, SEQ ID NO:1233, SEQ ID NO:1234, SEQID NO:1235,
 SEQ ID NO:1236, SEQ ID NO:1237, SEQ ID NO:1238, SEQ
 IDNO:1239, SEQ ID NO:1241, SEQ ID NO:1242, SEQ ID NO:
 1243, SEQ ID NO:1244, SEQ ID NO:1245, SEQ ID NO:1246,
 SEQ ID NO:1247, SEQ ID NO:1248, SEQ ID NO:1249, SEQ ID
 475 NO:1250, SEQ ID NO:1251, SEQ ID NO:1252, SEQ ID NO:1253,
 SEQ ID NO:1254, SEQ ID NO:1255, SEQ ID NO:1256, SEQ

IDNO:1257, SEQ ID NO:1259, SEQ ID NO:1260, SEQ ID NO:
 1261, SEQ ID NO:1262, SEQ ID NO:1263, SEQ ID NO:1264,
 SEQ ID NO:1265, SEQ ID NO:1266, SEQ ID NO:1267, SEQ ID
 480 NO:1268, SEQ ID NO:1269, SEQ ID NO:1270,SEQ ID NO:1271,
 SEQ ID NO:1272, SEQ ID NO:1273, SEQ ID NO:1275, SEQID
 NO:1276, SEQ ID NO:1277, SEQ ID NO:1278, SEQ ID NO:1279,
 SEQ ID NO:1280, SEQ ID NO:1281, SEQ ID NO:1282, SEQ ID
 NO:1283, SEQ ID NO:1284, SEQ ID NO:1285, SEQ ID NO:1286,
 485 SEQ ID NO:1287, SEQ ID NO:1289,SEQ ID NO:1290, SEQ ID
 NO:1291, SEQ ID NO:1292, SEQ ID NO:1293, SEQID NO:1294,
 SEQ ID NO:1295, SEQ ID NO:1296, SEQ ID NO:1297, SEQ
 IDNO:1298, SEQ ID NO:1299, SEQ ID NO:1300, SEQ ID NO:
 1301, SEQ ID NO:1303, SEQ ID NO:1304, SEQ ID NO:1305,
 490 SEQ ID NO:1306, SEQ ID NO:1307, SEQ ID NO:1308, SEQ ID
 NO:1310, SEQ ID NO:1311, SEQ ID NO:1312, SEQ ID NO:1313,
 SEQ ID NO:1314, SEQ ID NO:1315, SEQ ID NO:1316, SEQ
 IDNO:1317, SEQ ID NO:1318, SEQ ID NO:1319, SEQ ID NO:
 1320, SEQ ID NO:1322, SEQ ID NO:1323, SEQ ID NO:1324,
 495 SEQ ID NO:1325, SEQ ID NO:1326, SEQ ID NO:1327, SEQ ID
 NO:1328, SEQ ID NO:1330, SEQ ID NO:1331,SEQ ID NO:1332,
 SEQ ID NO:1333, SEQ ID NO:1334, SEQ ID NO:1335, SEQID
 NO:1336, SEQ ID NO:1337, SEQ ID NO:1339, SEQ ID NO:1340,
 SEQ ID NO:1341, SEQ ID NO:1342, SEQ ID NO:1343, SEQ ID
 500 NO:1344, SEQ ID NO:1345, SEQ ID NO:1346, SEQ ID NO:1347,
 SEQ ID NO:1349, SEQ ID NO:1350,SEQ ID NO:1351, SEQ ID
 NO:1352, SEQ ID NO:1353, SEQ ID NO:1354, SEQID NO:1355,
 SEQ ID NO:1356, SEQ ID NO:1357, SEQ ID NO:1358, SEQ
 IDNO:1360, SEQ ID NO:1361, SEQ ID NO:1362, SEQ ID NO:
 505 1363, SEQ ID NO:1364, SEQ ID NO:1365, SEQ ID NO:1367,
 SEQ ID NO:1368, SEQ ID NO:1369, SEQ ID NO:1370, SEQ ID
 NO:1371, SEQ ID NO:1375, SEQ ID NO:1376, SEQ ID NO:1377,
 SEQ ID NO:1378, SEQ ID NO:1379, SEQ ID NO:1381, SEQ
 IDNO:1382, SEQ ID NO:1383, SEQ ID NO:1384, SEQ ID NO:
 510 1385, SEQ ID NO:1387, SEQ ID NO:1388, SEQ ID NO:1389,

SEQ ID NO:1390, SEQ ID NO:1391, SEQ ID NO:1392, SEQ ID
 NO: 1393, SEQ ID NO:1395, SEQ ID NO:1396,SEQ ID NO:1397,
 SEQ ID NO:1398, SEQ ID NO:1399, SEQ ID NO:1400, SEQID
 NO:1402, SEQ ID NO:1403, SEQ ID NO:1404, SEQ ID NO:1405,
 515 SEQ ID NO:1406, SEQ ID NO:1407, SEQ ID NO:1409, SEQ ID
 NO:1410, SEQ ID NO:1412, SEQ ID NO:1413, SEQ ID NO:1414,
 SEQ ID NO:1415, SEQ ID NO:1416,SEQ ID NO:1417, SEQ ID
 NO:1419, SEQ ID NO:1420, SEQ ID NO:1421, SEQID NO:1422,
 SEQ ID NO:1423, SEQ ID NO:1424, SEQ ID NO:1425, SEQ
 520 IDNO:1427, SEQ ID NO:1428, SEQ ID NO:1429, SEQ ID NO:
 1430, SEQ ID NO:1431, SEQ ID NO:1432, SEQ ID NO:1433,
 SEQ ID NO:1434, SEQ ID NO:1435, SEQ ID NO:1437, SEQ ID
 NO:1438, SEQ ID NO:1439, SEQ ID NO:1440, SEQ ID NO:1441,
 SEQ ID NO:1442, SEQ ID NO:1444, SEQ ID NO:1445, SEQ
 525 IDNO:1446, SEQ ID NO:1447, SEQ ID NO:1448, SEQ ID NO:
 1449, SEQ ID NO:1451, SEQ ID NO:1452, SEQ ID NO:1453,
 SEQ ID NO:1454, SEQ ID NO:1455, SEQ ID NO:1456, SEQ ID
 NO:1458, SEQ ID NO:1459, SEQ ID NO:1461,SEQ ID NO:1462,
 SEQ ID NO:1463, SEQ ID NO:1464, SEQ ID NO:1465, SEQID
 530 NO:1466, SEQ ID NO:1468, SEQ ID NO:1469, SEQ ID NO:1470,
 SEQ ID NO:1472, SEQ ID NO:1474, SEQ ID NO:1475, SEQ ID
 NO:1476, SEQ ID NO:1477, SEQ ID NO:1479, SEQ ID NO:1480,
 SEQ ID NO:1481, SEQ ID NO:1482,SEQ ID NO:1483, SEQ ID
 NO:1484, SEQ ID NO:1485, SEQ ID NO:1486, SEQID NO:1488,
 535 SEQ ID NO:1490, SEQ ID NO:1491, SEQ ID NO:1492, SEQ
 IDNO:1493, SEQ ID NO:1495, SEQ ID NO:1496, SEQ ID NO:
 1497, SEQ ID NO:1498, SEQ ID NO:1500, SEQ ID NO:1502,
 SEQ ID NO:1503, SEQ ID NO:1504, SEQ ID NO:1505, SEQ ID
 NO:1507, SEQ ID NO:1509, SEQ ID NO:1512, SEQ ID NO:1513,
 540 SEQ ID NO:1514, SEQ ID NO:1515, SEQ ID NO:1517, SEQ
 IDNO:1518, SEQ ID NO:1519, SEQ ID NO:1521, SEQ ID NO:
 1522, SEQ ID NO:1523, SEQ ID NO:1524, SEQ ID NO:1525,
 SEQ ID NO:1527, SEQ ID NO:1528, SEQ ID NO:1529, SEQ ID
 NO:1530, SEQ ID NO:1531, SEQ ID NO:1533,SEQ ID NO:1534,

545 SEQ ID NO:1535, SEQ ID NO:1536, SEQ ID NO:1538, SEQ ID
 NO:1539, SEQ ID NO:1541, SEQ ID NO:1542, SEQ ID NO:1543,
 SEQ ID NO:1544, SEQ ID NO:1546, SEQ ID NO:1548, SEQ ID
 NO:1550, SEQ ID NO:1552, SEQ ID NO:1554, SEQ ID NO:1556,
 SEQ ID NO:1557, SEQ ID NO:1559, SEQ ID NO:1560, SEQ ID
 550 NO:1561, SEQ ID NO:1562, SEQ ID NO:1564, SEQ ID NO:1565,
 SEQ ID NO:1567, SEQ ID NO:1568, SEQ ID NO:1570, SEQ
 ID NO:1572, SEQ ID NO:1573, SEQ ID NO:1574, SEQ ID NO:
 1575, SEQ ID NO:1577, SEQ ID NO:1578, SEQ ID NO:1579,
 SEQ ID NO:1581, SEQ ID NO:1582, SEQ ID NO:1583, SEQ ID
 555 NO:1585, SEQ ID NO:1586, SEQ ID NO:1588, SEQ ID NO:1589,
 SEQ ID NO:1590, SEQ ID NO:1592, SEQ ID NO:1593, SEQ
 ID NO:1595, SEQ ID NO:1597, SEQ ID NO:1598, SEQ ID NO:
 1600, SEQ ID NO:1602, SEQ ID NO:1606, SEQ ID NO:1608,
 SEQ ID NO:1609, SEQ ID NO:1610, SEQ ID NO:1611, SEQ ID
 560 NO:1613, SEQ ID NO:1614, SEQ ID NO:1616, SEQ ID NO:1618,
 SEQ ID NO:1620, SEQ ID NO:1621, SEQ ID NO:1623, SEQ ID
 NO:1625, SEQ ID NO:1628, SEQ ID NO:1630, SEQ ID NO:1631,
 SEQ ID NO:1633, SEQ ID NO:1634, SEQ ID NO:1638, SEQ ID
 NO:1641, SEQ ID NO:1642, SEQ ID NO:1644, SEQ ID NO:1645,
 565 SEQ ID NO:1647, SEQ ID NO:1648, SEQ ID NO:1650, SEQ ID
 NO:1651, SEQ ID NO:1653, SEQ ID NO:1654, SEQ ID NO:1656,
 SEQ ID NO:1657, SEQ ID NO:1659, SEQ ID NO:1661, SEQ
 ID NO:1663, SEQ ID NO:1665, SEQ ID NO:1667, SEQ ID NO:
 1671, SEQ ID NO:1674, SEQ ID NO:1676, SEQ ID NO:1678,
 570 SEQ ID NO:1679, SEQ ID NO:1681, SEQ ID NO:1684, SEQ ID
 NO:1686, SEQ ID NO:1687, SEQ ID NO:1689, SEQ ID NO:1692,
 SEQ ID NO:1693, SEQ ID NO:1695, SEQ ID NO:1697, SEQ
 ID NO:1698, SEQ ID NO:1702, and SEQ ID NO:1703

, or (b) a polypeptide comprising an amino acid sequences
 575 in the amino acid sequences set forth in (a) in which several
 amino acids are deleted, replaced or added.

4. A polypeptide specific to enterohemorrhagic pathogenic *E.*
coli O-157:H7.

5. The polypeptide of claim 4 comprising

580 (a) an amino acid sequence selected from a group comprising the following SEQ IDs or a moiety thereof: SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14,

585 SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34,

590 SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54,

595 SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74,

600 SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94,

605 SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:

610 113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117, SEQ ID NO:118, SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:

124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID
NO:128, SEQ ID NO:129, SEQ ID NO:130, SEQ ID NO:131,
615 SEQ ID NO:133, SEQ ID NO:134, SEQ ID NO:135, SEQ ID NO:
136, SEQ ID NO:137, SEQ ID NO:138, SEQ ID NO:139, SEQ ID
NO:140, SEQ ID NO:141, SEQ ID NO:142, SEQ ID NO:143,
SEQ ID NO:144, SEQ ID NO:145, SEQ ID NO:146, SEQ ID NO:
147, SEQ ID NO:148, SEQ ID NO:149, SEQ ID NO:150, SEQ ID
620 NO:151, SEQ ID NO:152, SEQ ID NO:153, SEQ ID NO:154,
SEQ ID NO:155, SEQ ID NO:156, SEQ ID NO:157, SEQ ID NO:
158, SEQ ID NO:159, SEQ ID NO:160, SEQ ID NO:161, SEQ ID
NO:162, SEQ ID NO:163, SEQ ID NO:164, SEQ ID NO:165,
SEQ ID NO:166, SEQ ID NO:167, SEQ ID NO:168, SEQ ID NO:
625 169, SEQ ID NO:170, SEQ ID NO:171, SEQ ID NO:172, SEQ ID
NO:173, SEQ ID NO:174, SEQ ID NO:175, SEQ ID NO:176,
SEQ ID NO:177, SEQ ID NO:178, SEQ ID NO:179, SEQ ID NO:
180, SEQ ID NO:181, SEQ ID NO:182, SEQ ID NO:183, SEQ ID
NO:184, SEQ ID NO:185, SEQ ID NO:186, SEQ ID NO:187,
630 SEQ ID NO:188, SEQ ID NO:189, SEQ ID NO:190, SEQ ID NO:
191, SEQ ID NO:192, SEQ ID NO:193, SEQ ID NO:194, SEQ ID
NO:195, SEQ ID NO:196, SEQ ID NO:197, SEQ ID NO:198,
SEQ ID NO:199, SEQ ID NO:200, SEQ ID NO:201, SEQ ID NO:
202, SEQ ID NO:203, SEQ ID NO:204, SEQ ID NO:205, SEQ ID
635 NO:206, SEQ ID NO:207, SEQ ID NO:208, SEQ ID NO:209,
SEQ ID NO:210, SEQ ID NO:211, SEQ ID NO:212, SEQ ID NO:
213, SEQ ID NO:214, SEQ ID NO:215, SEQ ID NO:216, SEQ ID
NO:217, SEQ ID NO:218, SEQ ID NO:219, SEQ ID NO:220,
SEQ ID NO:221, SEQ ID NO:222, SEQ ID NO:223, SEQ ID NO:
640 224, SEQ ID NO:225, SEQ ID NO:226, SEQ ID NO:227, SEQ ID
NO:228, SEQ ID NO:229, SEQ ID NO:230, SEQ ID NO:231,
SEQ ID NO:232, SEQ ID NO:233, SEQ ID NO:234, SEQ ID NO:
235, SEQ ID NO:236, SEQ ID NO:237, SEQ ID NO:238, SEQ ID
NO:239, SEQ ID NO:240, SEQ ID NO:241, SEQ ID NO:242,
645 SEQ ID NO:243, SEQ ID NO:245, SEQ ID NO:246, SEQ ID NO:
247, SEQ ID NO:248, SEQ ID NO:249, SEQ ID NO:250, SEQ ID

NO:251, SEQ ID NO:252, SEQ ID NO:253, SEQ ID NO:254,
SEQ ID NO:255, SEQ ID NO:256, SEQ ID NO:257, SEQ ID NO:
258, SEQ ID NO:259, SEQ ID NO:260, SEQ ID NO:261, SEQ ID
650 NO:262, SEQ ID NO:263, SEQ ID NO:264, SEQ ID NO:265,
SEQ ID NO:266, SEQ ID NO:267, SEQ ID NO:268, SEQ ID NO:
269, SEQ ID NO:270, SEQ ID NO:271, SEQ ID NO:272, SEQ ID
NO:273, SEQ ID NO:274, SEQ ID NO:275, SEQ ID NO:276,
SEQ ID NO:277, SEQ ID NO:278, SEQ ID NO:279, SEQ ID NO:
655 280, SEQ ID NO:281, SEQ ID NO:282, SEQ ID NO:283, SEQ ID
NO:284, SEQ ID NO:285, SEQ ID NO:286, SEQ ID NO:287,
SEQ ID NO:288, SEQ ID NO:289, SEQ ID NO:290, SEQ ID NO:
291, SEQ ID NO:292, SEQ ID NO:293, SEQ ID NO:294, SEQ ID
NO:295, SEQ ID NO:296, SEQ ID NO:297, SEQ ID NO:298,
660 SEQ ID NO:299, SEQ ID NO:300, SEQ ID NO:301, SEQ ID NO:
302, SEQ ID NO:303, SEQ ID NO:304, SEQ ID NO:305, SEQ ID
NO:306, SEQ ID NO:307, SEQ ID NO:308, SEQ ID NO:309,
SEQ ID NO:310, SEQ ID NO:311, SEQ ID NO:312, SEQ ID NO:
313, SEQ ID NO:314, SEQ ID NO:315, SEQ ID NO:316, SEQ ID
665 NO:317, SEQ ID NO:318, SEQ ID NO:319, SEQ ID NO:320,
SEQ ID NO:321, SEQ ID NO:322, SEQ ID NO:323, SEQ ID NO:
324, SEQ ID NO:325, SEQ ID NO:326, SEQ ID NO:327, SEQ ID
NO:328, SEQ ID NO:329, SEQ ID NO:330, SEQ ID NO:331,
SEQ ID NO:332, SEQ ID NO:333, SEQ ID NO:334, SEQ ID NO:
670 335, SEQ ID NO:336, SEQ ID NO:338, SEQ ID NO:339, SEQ ID
NO:340, SEQ ID NO:341, SEQ ID NO:342, SEQ ID NO:343,
SEQ ID NO:344, SEQ ID NO:345, SEQ ID NO:346, SEQ ID NO:
347, SEQ ID NO:348, SEQ ID NO:349, SEQ ID NO:350, SEQ ID
NO:351, SEQ ID NO:352, SEQ ID NO:353, SEQ ID NO:354,
675 SEQ ID NO:355, SEQ ID NO:356, SEQ ID NO:357, SEQ ID NO:
358, SEQ ID NO:359, SEQ ID NO:360, SEQ ID NO:361, SEQ ID
NO:362, SEQ ID NO:363, SEQ ID NO:364, SEQ ID NO:365,
SEQ ID NO:366, SEQ ID NO:367, SEQ ID NO:368, SEQ ID NO:
369, SEQ ID NO:370, SEQ ID NO:371, SEQ ID NO:372, SEQ ID
680 NO:373, SEQ ID NO:374, SEQ ID NO:375, SEQ ID NO:376,

SEQ ID NO:377, SEQ ID NO:378, SEQ ID NO:379, SEQ ID NO:
380, SEQ ID NO:381, SEQ ID NO:382, SEQ ID NO:383, SEQ ID
NO:384, SEQ ID NO:385, SEQ ID NO:386, SEQ ID NO:387,
685 SEQ ID NO:388, SEQ ID NO:389, SEQ ID NO:390, SEQ ID NO:
391, SEQ ID NO:392, SEQ ID NO:393, SEQ ID NO:394, SEQ ID
NO:395, SEQ ID NO:396, SEQ ID NO:397, SEQ ID NO:398,
SEQ ID NO:399, SEQ ID NO:400, SEQ ID NO:401, SEQ ID NO:
402, SEQ ID NO:403, SEQ ID NO:404, SEQ ID NO:405, SEQ ID
NO:406, SEQ ID NO:407, SEQ ID NO:408, SEQ ID NO:409,
690 SEQ ID NO:411, SEQ ID NO:412, SEQ ID NO:413, SEQ ID NO:
414, SEQ ID NO:415, SEQ ID NO:416, SEQ ID NO:417, SEQ ID
NO:418, SEQ ID NO:419, SEQ ID NO:420, SEQ ID NO:421,
SEQ ID NO:422, SEQ ID NO:423, SEQ ID NO:424, SEQ ID NO:
425, SEQ ID NO:426, SEQ ID NO:427, SEQ ID NO:428, SEQ ID
695 NO:429, SEQ ID NO:430, SEQ ID NO:431, SEQ ID NO:432,
SEQ ID NO:433, SEQ ID NO:434, SEQ ID NO:435, SEQ ID NO:
436, SEQ ID NO:437, SEQ ID NO:438, SEQ ID NO:439, SEQ ID
NO:440, SEQ ID NO:441, SEQ ID NO:442, SEQ ID NO:443,
SEQ ID NO:444, SEQ ID NO:445, SEQ ID NO:446, SEQ ID NO:
700 447, SEQ ID NO:448, SEQ ID NO:449, SEQ ID NO:450, SEQ ID
NO:451, SEQ ID NO:452, SEQ ID NO:453, SEQ ID NO:454,
SEQ ID NO:455, SEQ ID NO:456, SEQ ID NO:457, SEQ ID NO:
458, SEQ ID NO:459, SEQ ID NO:460, SEQ ID NO:461, SEQ ID
NO:462, SEQ ID NO:463, SEQ ID NO:464, SEQ ID NO:465,
705 SEQ ID NO:466, SEQ ID NO:467, SEQ ID NO:468, SEQ ID NO:
469, SEQ ID NO:470, SEQ ID NO:471, SEQ ID NO:472, SEQ ID
NO:473, SEQ ID NO:474, SEQ ID NO:475, SEQ ID NO:476,
SEQ ID NO:477, SEQ ID NO:478, SEQ ID NO:479, SEQ ID NO:
480, SEQ ID NO:481, SEQ ID NO:482, SEQ ID NO:483, SEQ ID
710 NO:485, SEQ ID NO:486, SEQ ID NO:487, SEQ ID NO:488,
SEQ ID NO:489, SEQ ID NO:490, SEQ ID NO:491, SEQ ID NO:
492, SEQ ID NO:493, SEQ ID NO:494, SEQ ID NO:495, SEQ ID
NO:496, SEQ ID NO:497, SEQ ID NO:498, SEQ ID NO:499,
SEQ ID NO:500, SEQ ID NO:501, SEQ ID NO:502, SEQ ID NO:

715 503, SEQ ID NO:504, SEQ ID NO:505, SEQ ID NO:506, SEQ ID
NO:507, SEQ ID NO:508, SEQ ID NO:509, SEQ ID NO:510,
SEQ ID NO:511, SEQ ID NO:512, SEQ ID NO:513, SEQ ID NO:
514, SEQ ID NO:515, SEQ ID NO:516, SEQ ID NO:517, SEQ ID
NO:518, SEQ ID NO:519, SEQ ID NO:520, SEQ ID NO:521,
720 SEQ ID NO:522, SEQ ID NO:523, SEQ ID NO:524, SEQ ID NO:
525, SEQ ID NO:526, SEQ ID NO:527, SEQ ID NO:528, SEQ ID
NO:529, SEQ ID NO:530, SEQ ID NO:531, SEQ ID NO:532,
SEQ ID NO:533, SEQ ID NO:534, SEQ ID NO:535, SEQ ID NO:
536, SEQ ID NO:537, SEQ ID NO:538, SEQ ID NO:539, SEQ ID
725 NO:540, SEQ ID NO:541, SEQ ID NO:542, SEQ ID NO:543,
SEQ ID NO:544, SEQ ID NO:545, SEQ ID NO:546, SEQ ID NO:
547, SEQ ID NO:548, SEQ ID NO:549, SEQ ID NO:550, SEQ ID
NO:551, SEQ ID NO:552, SEQ ID NO:553, SEQ ID NO:555,
SEQ ID NO:556, SEQ ID NO:557, SEQ ID NO:558, SEQ ID NO:
730 559, SEQ ID NO:560, SEQ ID NO:561, SEQ ID NO:562, SEQ ID
NO:563, SEQ ID NO:564, SEQ ID NO:565, SEQ ID NO:566,
SEQ ID NO:567, SEQ ID NO:568, SEQ ID NO:569, SEQ ID NO:
570, SEQ ID NO:571, SEQ ID NO:572, SEQ ID NO:573, SEQ ID
NO:574, SEQ ID NO:575, SEQ ID NO:576, SEQ ID NO:577,
735 SEQ ID NO:578, SEQ ID NO:579, SEQ ID NO:580, SEQ ID NO:
581, SEQ ID NO:582, SEQ ID NO:583, SEQ ID NO:584, SEQ ID
NO:585, SEQ ID NO:586, SEQ ID NO:587, SEQ ID NO:588,
SEQ ID NO:589, SEQ ID NO:590, SEQ ID NO:591, SEQ ID NO:
592, SEQ ID NO:593, SEQ ID NO:594, SEQ ID NO:595, SEQ ID
740 NO:596, SEQ ID NO:597, SEQ ID NO:598, SEQ ID NO:599,
SEQ ID NO:600, SEQ ID NO:601, SEQ ID NO:602, SEQ ID NO:
603, SEQ ID NO:604, SEQ ID NO:605, SEQ ID NO:606, SEQ I
D NO:607, SEQ ID NO:608, SEQ ID NO:609, SEQ ID NO:610,
SEQ ID NO:611, SEQ ID NO:612, SEQ ID NO:613, SEQ ID NO:
745 614, SEQ ID NO:615, SEQ ID NO:616, SEQ ID NO:617, SEQ ID
NO:618, SEQ ID NO:619, SEQ ID NO:620, SEQ ID NO:621,
SEQ ID NO:622, SEQ ID NO:623, SEQ ID NO:624, SEQ ID NO:
625, SEQ ID NO:626, SEQ ID NO:627, SEQ ID NO:628, SEQ ID

NO:629, SEQ ID NO:631, SEQ ID NO:632, SEQ ID NO:633,
750 SEQ ID NO:634, SEQ ID NO:635, SEQ ID NO:636, SEQ ID NO:
637, SEQ ID NO:638, SEQ ID NO:639, SEQ ID NO:640, SEQ ID
NO:641, SEQ ID NO:642, SEQ ID NO:643, SEQ ID NO:644,
SEQ ID NO:645, SEQ ID NO:646, SEQ ID NO:647, SEQ ID NO:
648, SEQ ID NO:649, SEQ ID NO:650, SEQ ID NO:651, SEQ ID
755 NO:652, SEQ ID NO:653, SEQ ID NO:654, SEQ ID NO:655,
SEQ ID NO:656, SEQ ID NO:657, SEQ ID NO:658, SEQ ID NO:
659, SEQ ID NO:660, SEQ ID NO:661, SEQ ID NO:662, SEQ ID
NO:663, SEQ ID NO:664, SEQ ID NO:665, SEQ ID NO:666,
SEQ ID NO:667, SEQ ID NO:668, SEQ ID NO:669, SEQ ID NO:
760 670, SEQ ID NO:671, SEQ ID NO:672, SEQ ID NO:673, SEQ ID
NO:674, SEQ ID NO:675, SEQ ID NO:676, SEQ ID NO:677,
SEQ ID NO:678, SEQ ID NO:679, SEQ ID NO:680, SEQ ID NO:
681, SEQ ID NO:682, SEQ ID NO:683, SEQ ID NO:684, SEQ ID
NO:685, SEQ ID NO:686, SEQ ID NO:687, SEQ ID NO:688,
765 SEQ ID NO:690, SEQ ID NO:691, SEQ ID NO:692, SEQ ID NO:
693, SEQ ID NO:694, SEQ ID NO:695, SEQ ID NO:696, SEQ ID
NO:697, SEQ ID NO:698, SEQ ID NO:699, SEQ ID NO:700,
SEQ ID NO:701, SEQ ID NO:702, SEQ ID NO:703, SEQ ID NO:
704, SEQ ID NO:705, SEQ ID NO:706, SEQ ID NO:707, SEQ ID
770 NO:708, SEQ ID NO:709, SEQ ID NO:710, SEQ ID NO:711,
SEQ ID NO:712, SEQ ID NO:713, SEQ ID NO:714, SEQ ID NO:
715, SEQ ID NO:716, SEQ ID NO:717, SEQ ID NO:718, SEQ ID
NO:719, SEQ ID NO:720, SEQ ID NO:721, SEQ ID NO:722,
SEQ ID NO:723, SEQ ID NO:724, SEQ ID NO:725, SEQ ID NO:
775 726, SEQ ID NO:727, SEQ ID NO:728, SEQ ID NO:729, SEQ ID
NO:730, SEQ ID NO:731, SEQ ID NO:732, SEQ ID NO:733,
SEQ ID NO:734, SEQ ID NO:735, SEQ ID NO:736, SEQ ID NO:
737, SEQ ID NO:738, SEQ ID NO:739, SEQ ID NO:740, SEQ ID
NO:741, SEQ ID NO:742, SEQ ID NO:743, SEQ ID NO:744,
780 SEQ ID NO:745, SEQ ID NO:746, SEQ ID NO:747, SEQ ID NO:
748, SEQ ID NO:749, SEQ ID NO:750, SEQ ID NO:751, SEQ ID
NO:752, SEQ ID NO:753, SEQ ID NO:754, SEQ ID NO:756,

SEQ ID NO:757, SEQ ID NO:758, SEQ ID NO:759, SEQ ID NO:
760, SEQ ID NO:761, SEQ ID NO:762, SEQ ID NO:763, SEQ ID
785 NO:764, SEQ ID NO:765, SEQ ID NO:766, SEQ ID NO:767,
SEQ ID NO:768, SEQ ID NO:769, SEQ ID NO:770, SEQ ID NO:
771, SEQ ID NO:772, SEQ ID NO:773, SEQ ID NO:774, SEQ ID
NO:775, SEQ ID NO:776, SEQ ID NO:777, SEQ ID NO:778,
SEQ ID NO:779, SEQ ID NO:780, SEQ ID NO:781, SEQ ID NO:
790 782, SEQ ID NO:783, SEQ ID NO:784, SEQ ID NO:785, SEQ ID
NO:786, SEQ ID NO:787, SEQ ID NO:788, SEQ ID NO:789,
SEQ ID NO:790, SEQ ID NO:791, SEQ ID NO:792, SEQ ID NO:
793, SEQ ID NO:794, SEQ ID NO:795, SEQ ID NO:796, SEQ ID
NO:797, SEQ ID NO:798, SEQ ID NO:799, SEQ ID NO:800,
795 SEQ ID NO:801, SEQ ID NO:802, SEQ ID NO:803, SEQ ID NO:
804, SEQ ID NO:805, SEQ ID NO:806, SEQ ID NO:807, SEQ ID
NO:808, SEQ ID NO:809, SEQ ID NO:810, SEQ ID NO:811,
SEQ ID NO:812, SEQ ID NO:813, SEQ ID NO:814, SEQ ID NO:
815, SEQ ID NO:817, SEQ ID NO:818, SEQ ID NO:819, SEQ ID
800 NO:820, SEQ ID NO:821, SEQ ID NO:822, SEQ ID NO:823,
SEQ ID NO:824, SEQ ID NO:825, SEQ ID NO:826, SEQ ID NO:
827, SEQ ID NO:828, SEQ ID NO:829, SEQ ID NO:830, SEQ ID
NO:831, SEQ ID NO:832, SEQ ID NO:833, SEQ ID NO:834,
SEQ ID NO:835, SEQ ID NO:836, SEQ ID NO:837, SEQ ID NO:
805 838, SEQ ID NO:839, SEQ ID NO:840, SEQ ID NO:841, SEQ ID
NO:842, SEQ ID NO:843, SEQ ID NO:844, SEQ ID NO:845,
SEQ ID NO:846, SEQ ID NO:847, SEQ ID NO:848, SEQ ID NO:
849, SEQ ID NO:850, SEQ ID NO:851, SEQ ID NO:852, SEQ ID
NO:853, SEQ ID NO:854, SEQ ID NO:855, SEQ ID NO:856,
810 SEQ ID NO:857, SEQ ID NO:858, SEQ ID NO:859, SEQ ID NO:
860, SEQ ID NO:861, SEQ ID NO:862, SEQ ID NO:863, SEQ ID
NO:864, SEQ ID NO:865, SEQ ID NO:866, SEQ ID NO:867,
SEQ ID NO:868, SEQ ID NO:869, SEQ ID NO:870, SEQ ID NO:
871, SEQ ID NO:872, SEQ ID NO:873, SEQ ID NO:874. SEQ
815 ID NO:875, SEQ ID NO:877, SEQ ID NO:878, SEQ ID NO:879,
SEQ ID NO:880, SEQ ID NO:881, SEQ ID NO:882, SEQ ID NO:

883, SEQ ID NO:884, SEQ ID NO:885, SEQ ID NO:886, SEQ ID
NO:887, SEQ ID NO:888, SEQ ID NO:889, SEQ ID NO:890,
SEQ ID NO:891, SEQ ID NO:892, SEQ ID NO:893, SEQ ID NO:
820 894, SEQ ID NO:895, SEQ ID NO:896, SEQ ID NO:897, SEQ ID
NO:898, SEQ ID NO:899, SEQ ID NO:900, SEQ ID NO:901,
SEQ ID NO:902, SEQ ID NO:903, SEQ ID NO:904, SEQ ID NO:
905, SEQ ID NO:906, SEQ ID NO:907, SEQ ID NO:908, SEQ ID
NO:909, SEQ ID NO:910, SEQ ID NO:911, SEQ ID NO:912,
825 SEQ ID NO:913, SEQ ID NO:914, SEQ ID NO:915, SEQ ID NO:
916, SEQ ID NO:917, SEQ ID NO:918, SEQ ID NO:919, SEQ ID
NO:920, SEQ ID NO:921, SEQ ID NO:922, SEQ ID NO:923,
SEQ ID NO:924, SEQ ID NO:925, SEQ ID NO:926, SEQ ID NO:
928, SEQ ID NO:929, SEQ ID NO:930, SEQ ID NO:931, SEQ ID
830 NO:932, SEQ ID NO:933, SEQ ID NO:934, SEQ ID NO:935,
SEQ ID NO:936, SEQ ID NO:937, SEQ ID NO:938, SEQ ID NO:
939, SEQ ID NO:940, SEQ ID NO:941, SEQ ID NO:942, SEQ ID
NO:943, SEQ ID NO:944, SEQ ID NO:945, SEQ ID NO:946,
SEQ ID NO:947, SEQ ID NO:948, SEQ ID NO:949, SEQ ID NO:
835 950, SEQ ID NO:951, SEQ ID NO:952, SEQ ID NO:953, SEQ ID
NO:954, SEQ ID NO:955, SEQ ID NO:956, SEQ ID NO:957,
SEQ ID NO:958, SEQ ID NO:959, SEQ ID NO:960, SEQ ID NO:
961, SEQ ID NO:962, SEQ ID NO:963, SEQ ID NO:964, SEQ ID
NO:965, SEQ ID NO:966, SEQ ID NO:967, SEQ ID NO:968,
840 SEQ ID NO:969, SEQ ID NO:970, SEQ ID NO:971, SEQ ID NO:
972, SEQ ID NO:973, SEQ ID NO:974, SEQ ID NO:975, SEQ ID
NO:976, SEQ ID NO:977, SEQ ID NO:979, SEQ ID NO:980,
SEQ ID NO:981, SEQ ID NO:982, SEQ ID NO:983, SEQ ID NO:
984, SEQ ID NO:985, SEQ ID NO:986, SEQ ID NO:987, SEQ ID
845 NO:988, SEQ ID NO:989, SEQ ID NO:990, SEQ ID NO:991,
SEQ ID NO:992, SEQ ID NO:993, SEQ ID NO:994, SEQ ID NO:
995, SEQ ID NO:996, SEQ ID NO:997, SEQ ID NO:998, SEQ ID
NO:999, SEQ ID NO:1000, SEQ ID NO:1001, SEQ ID NO:1002,
SEQ ID NO:1003, SEQ ID NO:1004, SEQ ID NO:1005, SEQ ID
850 NO:1006, SEQ ID NO:1007, SEQ ID NO:1008, SEQ ID NO:1009,

SEQ ID NO:1010, SEQ ID NO:1011, SEQ ID NO:1012, SEQ ID
 NO:1014, SEQ ID NO:1015, SEQ ID NO:1016, SEQ ID NO:1017,
 SEQ ID NO:1018,SEQ ID NO:1019, SEQ ID NO:1020, SEQ ID
 NO:1021, SEQ ID NO:1022, SEQ ID NO:1023, SEQ ID NO:1024,
 855 SEQ ID NO:1025, SEQ ID NO:1026, SEQ ID NO:1027, SEQ ID
 NO:1028, SEQ ID NO:1030, SEQ ID NO:1031, SEQ ID NO:1032,
 SEQ ID NO:1033, SEQ ID NO:1034, SEQ ID NO:1035, SEQ ID
 NO:1036, SEQ ID NO:1037, SEQ ID NO:1038, SEQ ID NO:1039,
 SEQ ID NO:1040, SEQ ID NO:1041, SEQ ID NO:1042, SEQ ID
 860 NO:1043, SEQ ID NO:1044, SEQ ID NO:1045, SEQ ID NO:1046,
 SEQ ID NO:1047, SEQ ID NO:1048, SEQ ID NO:1049, SEQ ID
 NO:1050, SEQ ID NO:1051, SEQ ID NO:1052, SEQ ID NO:1053,
 SEQ ID NO:1054, SEQ ID NO:1056, SEQ ID NO:1057, SEQ ID
 NO:1058,SEQ ID NO:1059, SEQ ID NO:1061, SEQ ID NO:1062,
 865 SEQ ID NO:1063, SEQ ID NO:1064, SEQ ID NO:1065, SEQ ID
 NO:1066, SEQ ID NO:1067, SEQ ID NO:1068, SEQ ID NO:1069,
 SEQ ID NO:1070, SEQ ID NO:1071, SEQ ID NO:1072, SEQ ID
 NO:1073, SEQ ID NO:1074, SEQ ID NO:1075, SEQ ID NO:
 1076,SEQ ID NO:1077, SEQ ID NO:1078, SEQ ID NO:1079,
 870 SEQ ID NO:1080, SEQ ID NO:1081, SEQ ID NO:1082, SEQ ID
 NO:1083, SEQ ID NO:1084, SEQ ID NO:1085, SEQ ID NO:1086,
 SEQ ID NO:1087, SEQ ID NO:1088, SEQ ID NO:1089, SEQ ID
 NO:1090, SEQ ID NO:1091, SEQ ID NO:1092, SEQ ID NO:1094,
 SEQ ID NO:1095, SEQ ID NO:1096, SEQ ID NO:1097, SEQ ID
 875 NO:1098, SEQ ID NO:1099, SEQ ID NO:1100, SEQ ID NO:1101,
 SEQ ID NO:1102, SEQ ID NO:1103, SEQ ID NO:1104, SEQ ID
 NO:1105, SEQ ID NO:1106, SEQ ID NO:1107, SEQ ID NO:1108,
 SEQ ID NO:1109, SEQ ID NO:1110, SEQ ID NO:1111, SEQ ID
 NO:1112, SEQ ID NO:1113, SEQ ID NO:1114, SEQ ID NO:
 880 1115,SEQ ID NO:1116, SEQ ID NO:1117, SEQ ID NO:1118, SEQ
 ID NO:1119, SEQ ID NO:1120, SEQ ID NO:1121, SEQ ID NO:
 1122, SEQ ID NO:1123, SEQ ID NO:1124, SEQ ID NO:1125,
 SEQ ID NO:1126, SEQ ID NO:1127, SEQ ID NO:1129, SEQ ID
 NO:1130, SEQ ID NO:1131, SEQ ID NO:1132, SEQ ID NO:

885 1133,SEQ ID NO:1134, SEQ ID NO:1135, SEQ ID NO:1136,
 SEQ ID NO:1137, SEQID NO:1138, SEQ ID NO:1139, SEQ ID
 NO:1140, SEQ ID NO:1141, SEQ IDNO:1142, SEQ ID NO:1143,
 SEQ ID NO:1144, SEQ ID NO:1145, SEQ ID NO:1146, SEQ ID
 NO:1147, SEQ ID NO:1148, SEQ ID NO:1149, SEQ ID NO:1150,
 890 SEQ ID NO:1151, SEQ ID NO:1152, SEQ ID NO:1153, SEQ ID
 NO:1154, SEQ ID NO:1155, SEQ ID NO:1156, SEQ ID NO:1158,
 SEQ ID NO:1159, SEQ IDNO:1160, SEQ ID NO:1161, SEQ ID
 NO:1162, SEQ ID NO:1163, SEQ ID NO:1164, SEQ ID NO:1165,
 SEQ ID NO:1166, SEQ ID NO:1167, SEQ ID NO:1168, SEQ ID
 895 NO:1169, SEQ ID NO:1170, SEQ ID NO:1171, SEQ ID NO:
 1172,SEQ ID NO:1173, SEQ ID NO:1174, SEQ ID NO:1175,
 SEQ ID NO:1176, SEQID NO:1177, SEQ ID NO:1178, SEQ ID
 NO:1179, SEQ ID NO:1180, SEQ ID NO:1181, SEQ ID NO:1182,
 SEQ ID NO:1183, SEQ ID NO:1184, SEQ ID NO:1185, SEQ ID
 900 NO:1186, SEQ ID NO:1187, SEQ ID NO:1188, SEQ ID NO:
 1189,SEQ ID NO:1190, SEQ ID NO:1192, SEQ ID NO:1193,
 SEQ ID NO:1194, SEQID NO:1195, SEQ ID NO:1196, SEQ ID
 NO:1197, SEQ ID NO:1198, SEQ IDNO:1199, SEQ ID NO:1200,
 SEQ ID NO:1201, SEQ ID NO:1202, SEQ ID NO:1203, SEQ ID
 905 NO:1204, SEQ ID NO:1205, SEQ ID NO:1206, SEQ ID NO:1207,
 SEQ ID NO:1208, SEQ ID NO:1209, SEQ ID NO:1210, SEQ ID
 NO:1211, SEQ ID NO:1213, SEQ ID NO:1214, SEQ ID NO:1215,
 SEQ ID NO:1216, SEQ IDNO:1217, SEQ ID NO:1218, SEQ ID
 NO:1219, SEQ ID NO:1220, SEQ ID NO:1221, SEQ ID NO:1222,
 910 SEQ ID NO:1223, SEQ ID NO:1224, SEQ ID NO:1225, SEQ ID
 NO:1226, SEQ ID NO:1227, SEQ ID NO:1228, SEQ ID NO:
 1229,SEQ ID NO:1230, SEQ ID NO:1231, SEQ ID NO:1232,
 SEQ ID NO:1233, SEQID NO:1234, SEQ ID NO:1235, SEQ ID
 NO:1236, SEQ ID NO:1237, SEQ ID NO:1238, SEQ ID NO:1239,
 915 SEQ ID NO:1241, SEQ ID NO:1242, SEQ ID NO:1243, SEQ ID
 NO:1244, SEQ ID NO:1245, SEQ ID NO:1246, SEQ ID NO:
 1247,SEQ ID NO:1248, SEQ ID NO:1249, SEQ ID NO:1250,
 SEQ ID NO:1251, SEQID NO:1252, SEQ ID NO:1253, SEQ ID

NO:1254, SEQ ID NO:1255, SEQ IDNO:1256, SEQ ID NO:1257,
920 SEQ ID NO:1259, SEQ ID NO:1260, SEQ ID NO:1261, SEQ ID
NO:1262, SEQ ID NO:1263, SEQ ID NO:1264, SEQ ID NO:1265,
SEQ ID NO:1266, SEQ ID NO:1267, SEQ ID NO:1268, SEQ ID
NO:1269, SEQ ID NO:1270, SEQ ID NO:1271, SEQ ID NO:1272,
SEQ ID NO:1273, SEQ IDNO:1275, SEQ ID NO:1276, SEQ ID
925 NO:1277, SEQ ID NO:1278, SEQ ID NO:1279, SEQ ID NO:1280,
SEQ ID NO:1281, SEQ ID NO:1282, SEQ ID NO:1283, SEQ ID
NO:1284, SEQ ID NO:1285, SEQ ID NO:1286, SEQ ID NO:
1287,SEQ ID NO:1289, SEQ ID NO:1290, SEQ ID NO:1291,
SEQ ID NO:1292, SEQID NO:1293, SEQ ID NO:1294, SEQ ID
930 NO:1295, SEQ ID NO:1296, SEQ ID NO:1297, SEQ ID NO:1298,
SEQ ID NO:1299, SEQ ID NO:1300, SEQ ID NO:1301, SEQ ID
NO:1303, SEQ ID NO:1304, SEQ ID NO:1305, SEQ ID NO:
1306,SEQ ID NO:1307, SEQ ID NO:1308, SEQ ID NO:1310,
SEQ ID NO:1311, SEQID NO:1312, SEQ ID NO:1313, SEQ ID
935 NO:1314, SEQ ID NO:1315, SEQ IDNO:1316, SEQ ID NO:1317,
SEQ ID NO:1318, SEQ ID NO:1319, SEQ ID NO:1320, SEQ ID
NO:1322, SEQ ID NO:1323, SEQ ID NO:1324, SEQ ID NO:1325,
SEQ ID NO:1326, SEQ ID NO:1327, SEQ ID NO:1328, SEQ ID
NO:1330, SEQ ID NO:1331, SEQ ID NO:1332, SEQ ID NO:1333,
940 SEQ ID NO:1334, SEQ IDNO:1335, SEQ ID NO:1336, SEQ ID
NO:1337, SEQ ID NO:1339, SEQ ID NO:1340, SEQ ID NO:1341,
SEQ ID NO:1342, SEQ ID NO:1343, SEQ ID NO:1344, SEQ ID
NO:1345, SEQ ID NO:1346, SEQ ID NO:1347, SEQ ID NO:
1349,SEQ ID NO:1350, SEQ ID NO:1351, SEQ ID NO:1352,
945 SEQ ID NO:1353, SEQID NO:1354, SEQ ID NO:1355, SEQ ID
NO:1356, SEQ ID NO:1357, SEQ ID NO:1358, SEQ ID NO:1360,
SEQ ID NO:1361, SEQ ID NO:1362, SEQ ID NO:1363, SEQ ID
NO:1364, SEQ ID NO:1365, SEQ ID NO:1367, SEQ ID NO:
1368,SEQ ID NO:1369, SEQ ID NO:1370, SEQ ID NO:1371,
950 SEQ ID NO:1375, SEQID NO:1376, SEQ ID NO:1377, SEQ ID
NO:1378, SEQ ID NO:1379, SEQ IDNO:1381, SEQ ID NO:1382,
SEQ ID NO:1383, SEQ ID NO:1384, SEQ ID NO:1385, SEQ ID

NO:1387, SEQ ID NO:1388, SEQ ID NO:1389, SEQ ID NO:1390,
 SEQ ID NO:1391, SEQ ID NO:1392, SEQ ID NO:1393, SEQ ID
 955 NO:1395, SEQ ID NO:1396, SEQ ID NO:1397, SEQ ID NO:1398,
 SEQ ID NO:1399, SEQ IDNO:1400, SEQ ID NO:1402, SEQ ID
 NO:1403, SEQ ID NO:1404, SEQ ID NO:1405, SEQ ID NO:1406,
 SEQ ID NO:1407, SEQ ID NO:1409, SEQ ID NO:1410, SEQ ID
 NO:1412, SEQ ID NO:1413, SEQ ID NO:1414, SEQ ID NO:
 960 1415,SEQ ID NO:1416, SEQ ID NO:1417, SEQ ID NO:1419,
 SEQ ID NO:1420, SEQID NO:1421, SEQ ID NO:1422, SEQ ID
 NO:1423, SEQ ID NO:1424, SEQ ID NO:1425, SEQ ID NO:1427,
 SEQ ID NO:1428, SEQ ID NO:1429, SEQ ID NO:1430, SEQ ID
 NO:1431, SEQ ID NO:1432, SEQ ID NO:1433, SEQ ID NO:
 965 1434,SEQ ID NO:1435, SEQ ID NO:1437, SEQ ID NO:1438,
 SEQ ID NO:1439, SEQID NO:1440, SEQ ID NO:1441, SEQ ID
 NO:1442, SEQ ID NO:1444, SEQ IDNO:1445, SEQ ID NO:1446,
 SEQ ID NO:1447, SEQ ID NO:1448, SEQ ID NO:1449, SEQ ID
 NO:1451, SEQ ID NO:1452, SEQ ID NO:1453, SEQ ID NO:1454,
 970 SEQ ID NO:1455, SEQ ID NO:1456, SEQ ID NO:1458, SEQ ID
 NO:1459, SEQ ID NO:1461, SEQ ID NO:1462, SEQ ID NO:1463,
 SEQ ID NO:1464, SEQ IDNO:1465, SEQ ID NO:1466, SEQ ID
 NO:1468, SEQ ID NO:1469, SEQ ID NO:1470, SEQ ID NO:1472,
 SEQ ID NO:1474, SEQ ID NO:1475, SEQ ID NO:1476, SEQ ID
 975 NO:1477, SEQ ID NO:1479, SEQ ID NO:1480, SEQ ID NO:
 1481,SEQ ID NO:1482, SEQ ID NO:1483, SEQ ID NO:1484,
 SEQ ID NO:1485, SEQID NO:1486, SEQ ID NO:1488, SEQ ID
 NO:1490, SEQ ID NO:1491, SEQ ID NO:1492, SEQ ID NO:1493,
 SEQ ID NO:1495, SEQ ID NO:1496, SEQ ID NO:1497, SEQ ID
 980 NO:1498, SEQ ID NO:1500, SEQ ID NO:1502, SEQ ID NO:
 1503,SEQ ID NO:1504, SEQ ID NO:1505, SEQ ID NO:1507,
 SEQ ID NO:1509, SEQID NO:1512, SEQ ID NO:1513, SEQ ID
 NO:1514, SEQ ID NO:1515, SEQ IDNO:1517, SEQ ID NO:1518,
 SEQ ID NO:1519, SEQ ID NO:1521, SEQ ID NO:1522, SEQ ID
 985 NO:1523, SEQ ID NO:1524, SEQ ID NO:1525, SEQ ID NO:1527,
 SEQ ID NO:1528, SEQ ID NO:1529, SEQ ID NO:1530, SEQ ID

NO:1531, SEQ ID NO:1533, SEQ ID NO:1534, SEQ ID NO:1535,
SEQ ID NO:1536, SEQ IDNO:1538, SEQ ID NO:1539, SEQ ID
NO:1541, SEQ ID NO:1542, SEQ ID NO:1543, SEQ ID NO:1544,
990 SEQ ID NO:1546, SEQ ID NO:1548, SEQ ID NO:1550, SEQ ID
NO:1552, SEQ ID NO:1554, SEQ ID NO:1556, SEQ ID NO:
1557,SEQ ID NO:1559, SEQ ID NO:1560, SEQ ID NO:1561,
SEQ ID NO:1562, SEQID NO:1564, SEQ ID NO:1565, SEQ ID
NO:1567, SEQ ID NO:1568, SEQ ID NO:1570, SEQ ID NO:1572,
995 SEQ ID NO:1573, SEQ ID NO:1574, SEQ ID NO:1575, SEQ ID
NO:1577, SEQ ID NO:1578, SEQ ID NO:1579, SEQ ID NO:
1581,SEQ ID NO:1582, SEQ ID NO:1583, SEQ ID NO:1585,
SEQ ID NO:1586, SEQID NO:1588, SEQ ID NO:1589, SEQ ID
NO:1590, SEQ ID NO:1592, SEQ IDNO:1593, SEQ ID NO:1595,
1000 SEQ ID NO:1597, SEQ ID NO:1598, SEQ ID NO:1600, SEQ ID
NO:1602, SEQ ID NO:1606, SEQ ID NO:1608, SEQ ID NO:1609,
SEQ ID NO:1610, SEQ ID NO:1611, SEQ ID NO:1613, SEQ ID
NO:1614, SEQ ID NO:1616, SEQ ID NO:1618, SEQ ID NO:1620,
SEQ ID NO:1621, SEQ IDNO:1623, SEQ ID NO:1625, SEQ ID
1005 NO:1628, SEQ ID NO:1630, SEQ ID NO:1631, SEQ ID NO:1633,
SEQ ID NO:1634, SEQ ID NO:1638, SEQ ID NO:1641, SEQ ID
NO:1642, SEQ ID NO:1644, SEQ ID NO:1645, SEQ ID NO:
1647,SEQ ID NO:1648, SEQ ID NO:1650, SEQ ID NO:1651,
SEQ ID NO:1653, SEQID NO:1654, SEQ ID NO:1656, SEQ ID
1010 NO:1657, SEQ ID NO:1659, SEQ ID NO:1661, SEQ ID NO:1663,
SEQ ID NO:1665, SEQ ID NO:1667, SEQ ID NO:1671, SEQ ID
NO:1674, SEQ ID NO:1676, SEQ ID NO:1678, SEQ ID NO:
1679,SEQ ID NO:1681, SEQ ID NO:1684, SEQ ID NO:1686,
SEQ ID NO:1687, SEQID NO:1689, SEQ ID NO:1692, SEQ ID
1015 NO:1693, SEQ ID NO:1695, SEQ IDNO:1697, SEQ ID NO:1698,
SEQ ID NO:1702, and SEQ ID NO:1703

, or (b) an amino acid sequence in the amino acid
sequences set forth in (a) in which several amino acids are
deleted, replaced or added.

1020 6. A vector containing the nucleic-acid molecule of claim 1

as an inserted substance.

7. The vector of claim 6, wherein the inserted substance is linked with an element of transcriptional regulation in their action.

1025 8. A host cell which is transformed with the vector of claim 7.

9. A method of producing a polypeptide specific to O-157:H7 comprising cultivation of the host cell of claim 8.

1030 10. An oligonucleotide or polynucleotide specific to enterohemorrhagic pathogenic-E. coli O-157:H7 comprising a nucleotide sequence constituted of at least 8 nucleotides in

(a) a nucleotide sequence selected from a group comprising the following SEQ IDs: SEQ ID NO:1, SEQ ID NO:
1035 132, SEQ ID NO:244,SEQ ID NO:337, SEQ ID NO:410, SEQ ID NO:484, SEQ ID NO:554, SEQ ID NO:630, SEQ ID NO:689, SEQ ID NO:755, SEQ ID NO:816, SEQ ID NO:876,SEQ ID NO:927, SEQ ID NO:978, SEQ ID NO:1013, SEQ ID NO:1029, SEQ IDNO:1055, SEQ ID NO:1060, SEQ ID NO:1093, SEQ ID NO:
1040 1128, SEQ ID NO:1157, SEQ ID NO:1191, SEQ ID NO:1212, SEQ ID NO:1240, SEQ ID NO:1258, SEQ ID NO:1274, SEQ ID NO:1288, SEQ ID NO:1302, SEQ ID NO:1309,SEQ ID NO:1321, SEQ ID NO:1329, SEQ ID NO:1338, SEQ ID NO:1348, SEQID NO:1359, SEQ ID NO:1366, SEQ ID NO:1374, SEQ ID NO:1380,
1045 SEQ ID NO:1386, SEQ ID NO:1394, SEQ ID NO:1401, SEQ ID NO:1408, SEQ ID NO:1411, SEQ ID NO:1418, SEQ ID NO:1426, SEQ ID NO:1436, SEQ ID NO:1443,SEQ ID NO:1450, SEQ ID NO:1457, SEQ ID NO:1460, SEQ ID NO:1467, SEQID NO:1471, SEQ ID NO:1473, SEQ ID NO:1478, SEQ ID NO:1487, SEQ
1050 IDNO:1489, SEQ ID NO:1494, SEQ ID NO:1499, SEQ, ID NO:1501, SEQ ID NO:1506, SEQ ID NO:1508, SEQ ID NO:1510, SEQ ID NO:1511, SEQ ID NO:1516, SEQ ID NO:1520, SEQ ID NO:1526, SEQ ID NO:1532, SEQ ID NO:1537,SEQ ID NO:1540, SEQ ID NO:1545, SEQ ID NO:1547, SEQ ID NO:1549, SEQID

1055 NO:1551, SEQ ID NO:1553, SEQ ID NO:1555, SEQ ID NO:1558,
 SEQ ID NO:1563, SEQ ID NO:1566, SEQ ID NO:1569, SEQ ID
 NO:1571, SEQ ID NO:1576, SEQ ID NO:1580, SEQ ID NO:1584,
 SEQ ID NO:1587, SEQ ID NO:1591,SEQ ID NO:1594, SEQ ID
 NO:1596, SEQ ID NO:1599, SEQ ID NO:1601, SEQID NO:1603,
 1060 SEQ ID NO:1604, SEQ ID NO:1605, SEQ ID NO:1607, SEQ
 IDNO:1612, SEQ ID NO:1615, SEQ ID NO:1617, SEQ ID NO:
 1619, SEQ ID NO:1622, SEQ ID NO:1624, SEQ ID NO:1626,
 SEQ ID NO:1627, SEQ ID NO:1629, SEQ ID NO:1632, SEQ ID
 NO:1635, SEQ ID NO:1636, SEQ ID NO:1637, SEQ ID NO:1639,
 1065 SEQ ID NO:1640, SEQ ID NO:1643, SEQ ID NO:1646, SEQ
 IDNO:1649, SEQ ID NO:1652, SEQ ID NO:1655, SEQ ID NO:
 1658, SEQ ID NO:1660, SEQ ID NO:1662, SEQ ID NO:1664,
 SEQ ID NO:1666, SEQ ID NO:1668, SEQ ID NO:1669, SEQ ID
 NO:1670, SEQ ID NO:1672, SEQ ID NO:1673,SEQ ID NO:1675,
 1070 SEQ ID NO:1677, SEQ ID NO:1680, SEQ ID NO:1682, SEQID
 NO:1683, SEQ ID NO:1685, SEQ ID NO:1688, SEQ ID NO:1690,
 SEQ ID NO:1691, SEQ ID NO:1694, SEQ ID NO:1696, SEQ ID
 NO:1699, SEQ ID NO:1700, SEQ ID NO:1701, SEQ ID NO:1704,
 SEQ ID NO:1705, SEQ ID NO:1706,SEQ ID NO:1707, SEQ ID
 1075 NO:1708, SEQ ID NO:1709, SEQ ID NO:1710, SEQID NO:1711,
 SEQ ID NO:1712, SEQ ID NO:1713, SEQ ID NO:1715, SEQ
 IDNO:1716, SEQ ID NO:1717, SEQ ID NO:1718,, SEQ ID NO:
 1719, SEQ ID NO:1720, SEQ ID NO:1721, SEQ ID NO:1722,
 SEQ ID NO:1723, SEQ ID NO:1724, SEQ ID NO:1725, SEQ ID
 1080 NO:1726, SEQ ID NO:1727, SEQ ID NO:1728,SEQ ID NO:1729,
 SEQ ID NO:1730, SEQ ID NO:1731, SEQ ID NO:1732, SEQID
 NO:1733, SEQ ID NO:1734, SEQ ID NO:1735, SEQ ID NO:1736,
 SEQ ID NO:1737, SEQ ID NO:1738, SEQ ID NO:1739, SEQ ID
 NO:1740, SEQ ID NO:1741, SEQ ID NO:1742, SEQ ID NO:1743,
 1085 SEQ ID NO:1744, SEQ ID NO:1745,SEQ ID NO:1746, SEQ ID
 NO:1747, SEQ ID NO:1748, SEQ ID NO:1749, SEQID NO:1750,
 SEQ ID NO:1751, SEQ ID NO:1752, SEQ ID NO:1753, SEQ
 IDNO:1754, SEQ ID NO:1755, SEQ ID NO:1756, SEQ ID NO:

1757, SEQ ID NO:1758, SEQ ID NO:1759, SEQ ID NO:1760,
1090 SEQ ID NO:1761, SEQ ID NO:1762, SEQ ID NO:1763, SEQ ID
NO:1764, SEQ ID NO:1765, SEQ ID NO:1766, SEQ ID NO:1767,
SEQ ID NO:1768, SEQ ID NO:1769, SEQ ID NO:1770, SEQ
IDNO:1771, SEQ ID NO:1772, SEQ ID NO:1773, SEQ ID NO:
1774, SEQ ID NO:1775, SEQ ID NO:1776, SEQ ID NO:1777,
1095 SEQ ID NO:1778, SEQ ID NO:1779, SEQ ID NO:1780, SEQ ID
NO:1781, SEQ ID NO:1782, SEQ ID NO:1783,SEQ ID NO:1784,
SEQ ID NO:1785, SEQ ID NO:1786, SEQ ID NO:1787, SEQID
NO:1788, SEQ ID NO:1789, SEQ ID NO:1790, SEQ ID NO:1791,
SEQ ID NO:1792, SEQ ID NO:1793, SEQ ID NO:1794, SEQ ID
1100 NO:1795, SEQ ID NO:1796, SEQ ID NO:1797, SEQ ID NO:1798,
SEQ ID NO:1799, SEQ ID NO:1800,SEQ ID NO:1801, SEQ ID
NO:1802, SEQ ID NO:1803, SEQ ID NO:1804, SEQID NO:1805,
SEQ ID NO:1806, SEQ ID NO:1807, SEQ ID NO:1808, SEQ
IDNO:1809, SEQ ID NO:1810, SEQ ID NO:1811, SEQ ID NO:
1105 1812, SEQ ID NO:1813, SEQ ID NO:1814, SEQ ID NO:1815,
SEQ ID NO:1816, SEQ ID NO:1817, SEQ ID NO:1818, SEQ ID
NO:1819, SEQ ID NO:1820, SEQ ID NO:1821, SEQ ID NO:1822,
SEQ ID NO:1823, SEQ ID NO:1824, SEQ ID NO:1825, SEQ
IDNO:1826, SEQ ID NO:1827, SEQ ID NO:1828, SEQ ID NO:
1110 1829, SEQ ID NO:1830, SEQ ID NO:1831, SEQ ID NO:1832,
SEQ ID NO:1833, SEQ ID NO:1834, SEQ ID NO:1835, SEQ ID
NO:1836, SEQ ID NO:1837, SEQ ID NO:1838,SEQ ID NO:1839,
SEQ ID NO:1840, SEQ ID NO:1841, SEQ ID NO:1842, SEQID
NO:1843, SEQ ID NO:1844, SEQ ID NO:1845, SEQ ID NO:1846.
1115 SEQ ID NO:1847, SEQ ID NO:1848, SEQ ID NO:1849, SEQ ID
NO:1850, SEQ ID NO:1851, SEQ ID NO:1852, SEQ ID NO:1853,
SEQ ID NO:1854, SEQ ID NO:1855,SEQ ID NO:1856, SEQ ID
NO:1857, SEQ ID NO:1858, SEQ ID NO:1859, SEQID NO:1860,
SEQ ID NO:1861, SEQ ID NO:1862, SEQ ID NO:1863, SEQ
1120 IDNO:1864, SEQ ID NO:1865, and SEQ ID NO:1866

,and/or (b) a complementary nucleotide sequence to
the nucleic-acid sequence set forth in (a).

11. Use of the oligonucleotide or polynucleotide of claim 10 as a probe for hybridization or a primer for PCR.
- 1125 12. An use of the oligonucleotide or polynucleotide of claim 11 for detection or diagnosis of O-157 infection.
13. A vaccine composition comprising the nucleic-acid molecule of claim 1 or its fragment, or the oligonucleotide or polynucleotide of claim 10 and a pharmaceutically acceptable carrier.
- 1130 14. A vaccine composition comprising the polypeptide of claim 4 or its fragment and a pharmaceutically acceptable carrier.
15. An antibody molecule specifically recognizing the polypeptide of claim 4.
- 1135 16. A DNA microarray or DNA chip including the nucleic-acid molecule of claim 1 and/or at least one of the oligonucleotide or polynucleotide of claim 10.
17. Use of the DNA microarray or DNA chip for deteciton of O-157 infection or classification of O-157.
- 1140 18. A method of screening a compound useful for prevention or therapy of O-157 infection and a symptom caused thereby, using the nucleic-acid molecule of claim 1 or fragment thereof, or the polypeptide of claim 4 or fragment thereof.

1145

DESCRIPTION

A nucleic-acid molecule and a polypeptide specific to
enterohemorrhagic

E. coli O-157:H7 and a method of using thereof

1150 [0001]

INDUSTRIAL APPLICABLE FIELDS

The present invention relates to a novel nucleic-acid molecule and a polypeptide specific to O-157:H7 as well as use thereof.

1155 [0002]

BACKGROUND ART

Although *E. coli* also inhabits large intestine of healthy human, most *E. coli* especially causes no disease. However a part of *E. coli* infects the intestine of human to cause food poisoning such as enterogastritis and diarrhea. These are referred to as pathogenic *E. coli* and classified mainly into the following 5 categories: Enterotoxigenic *Escherichia coli*: ETEC, Enteroinvasive *Escherichia coli*: EIEC, Enteropathogenic *Escherichia coli*: EPEC, Enterohemorrhagic *Escherichia coli*: EHEC, Enteroadherent *Escherichia coli* ; EAEC

1160

1165

[0003]

EHEC therein includes *E. coli* which cause, as a main symptom, severe abdominal pain, diarrhea and/or hematochezia, in especially a child and an aged person, a serious complication such as renal dysfunction and haemolytic uraemic syndrome (HUS) and, in some cases, lead a patient to death. A main pathogenic bacterium therein is O-157:H7 (hereinafter referred to as "O-157"). O-157 belongs to a serotype different from that of EPEC or enteroinvasive *E. coli* which has been reported. In addition, it has been reported as a pathogenic *E. coli* which produces no thermolabile enterotoxin (LT) and thermostable enterotoxin (ST) by Riley et al. (Riley LW, et al., N. Engl. J. Med. 308 (1983), p.681-685). Furthermore, O-157 and EHEC

1170

1175

are also referred to as Verotoxin-producing *Escherichia coli*
1180 (VTEC), since it has been revealed that extracellular toxin
produced by them is Verotoxin (VT).
[0004]

The verotoxin (VT) produced by EHEC (or VTEC) is
identified as toxin which has potent cytotoxicity on Vero cells,
1185 African green monkey kidney cells. O'Brien et al. (J. Infect
Dis. 146 (1982), p.763-769) reported that its toxicity was
neutralized by an antibody to Shiga toxin produced by
dysentery bacillus, and referred the toxin to as Shiga-like toxin.
The verotoxin includes two major types (VT1 and VT2). Since
1190 the verotoxins are similar to Shiga toxin, they are also referred
to as SLT1 (Shiga-Like Toxin 1) and SLT2 respectively. VT1 is
identical to Shiga toxin, or different in 1 amino acid merely.
VT2 has homology of approximately 56% at amino acid level to
VT1 (Jackson M.P. et al., FEMS Microbial Lett. 44 (1987) p.
1195 109-114), whereas their antigenicity are little common. The
verotoxin and the Shiga toxin has the same N-glycosidic
activity as that of lysin which is a potent phytotoxin derived
from a plant. Their effects and functions are for inhibiting
linkage of an aminoacyl tRNA to a ribosome to inhibit protein
1200 synthesis by hydrolyzing an N-glycosidic linkage at an
adenosine in 28S ribosomal RNA constituting mammalian
eukaryotic 60S ribosome, thereby resulting in cell death.
Especially, the verotoxin cause damage to a vascular
endothelial cell such as large intestine and a renal tubular cell
1205 to cause haemolytic uraemic syndrome and the like.
[0005]

As mentioned above, O-157 causes hemorrhagic colitis
and sometimes complicates haemolytic uraemic syndrome or
encephalopathy which expose patient's life to danger. Up to
1210 now, none of effective methods for inhibiting or preventing
progression to haemolytic uraemic syndrome have been
established. In addition, administration of an antibacterial

agent such as antibiotic promote the extracellular release of VT, sometimes resulting in making the symptom worse. Therefore,
1215 definitive diagnosis of infection is important at early stage of the infection.

[0006]

Several methods are known as methods for diagnosis of the O-157 infection, i.e. the methods for distinguishing O-157
1220 from nonpathogenic or other pathogenic *E. coli*. One of them applies a feature that O-157:H7 is different from general *E. coli* and other known EPEC in the point that O-157:H7 produces no β -glucuronidase and ferments no sorbitol of saccharide, or do after some delay. This method has been used widely. However,
1225 these methods have the weak point of taking time and lacking rapidity. Further, although the presence of O-157 capable of degrading sorbitol is reported, these methods can not detect such bacterium. On the other hand, reversed passive latex aggregate reaction using an antibody to lipopolysaccharide
1230 antigen of O-157 or an antibody to the verotoxin is known. These methods can detect the bacterium producing VT rapidly and conveniently, but their detection sensitivity is not sufficient. Especially as to verotoxin, bacteria producing the toxin are not restricted to O-157, thus these methods have a
1235 task [should be solved] as methods for detecting O-157.

[0007]

Further, molecular biological methods, specifically, hybridization assay and PCR assay, are performed as the methods for detecting O-157. Especially, PCR is of extremely
1240 high detection sensitivity, high rapidity and high convenience, resulting in increasing use of it in recent years. Main terget of PCT etc. is VT gene of VTEC such as O-157. However, as mentioned above, *E. coli* other than O-157 also has the VT gene, and furthermore, multiple mutants of VT gene are known, thus
1245 there is a task [should be solved] as definitive methods for diagnosis of O-157. Moreover, although pulsed-field gel

electrophoresis (PFGE) is use for detection of O-157, an apparatus required for performing this method is expensive, and the method requires long time for detection and
1250 considerably skilled technique. In addition, the number of strains which can be analysed at once is limited and comparison of data of O-157 at different institutions is not easy. Therefore, there is need for a method which is of rapid, convenient, high detection sensitivity, high confidence and ease of comparison
1255 and exchange of data between different institutions.
[0008]

On the other hand, although antibacterial agents considered to be effective to O-157, such as antibiotic, are known, the presence of drug-resistant bacteria has also been
1260 reported. In addition, as mentioned above, VT is released to extracellular space by administration of antibiotics, sometimes resulting in making the patient's symptom worse. Therefore, there is a requirement for development of a method different from the method for therapy of infectious disease caused by
1265 O-157 using these antibacterial agents, a method for therapy and/or prevention of the symptom caused by VT, and detailed genetic information of O-157 which may serve as a guidance thereto.
[0009]

1270 PROBLEMS TO BE SOLVED BY THE INVENTION

Accordingly, the task of the present invention is providing a nucleic-acid molecule, a polypeptide, genetic information thereof and a method of using them which may be useful for detection and therapy of enterohemorrhagic pathogenic E. coli
1275 O-157:H7 infection.
[0010]

Means To Solve The Problem

We have found genetic information specific to O-157:H7 which is not present in other E. coli including nonpathogenic E.
1280 coli by analyzing whole genetic information of

enterohemorrhagic pathogenic-*E. coli* O-157:H7 Sakai (RIMD 0509952). Therefore, the present invention relates to the genetic information specific to O-157:H7 and the use thereof. The genetic information includes, but not restricted to, a
1285 nucleotide sequence on genome, a gene, a polypeptide encoded thereby, an amino acid sequence thereof and the like.
[0011]

Therefore, the present invention relates to a nucleic-acid molecule specific to enterohemorrhagic pathogenic-*E. coli*
1290 O-157:H7. In a preferred embodiment, the present invention relates to a nucleic-acid molecule having

(a) a nucleotide sequence selected from a group comprising the following SEQ IDs: SEQ ID NO:1, SEQ ID NO:132, SEQ ID NO:244, SEQ ID NO:337, SEQ ID NO:410, SEQ ID
1295 NO:484, SEQ ID NO:554, SEQ ID NO:630, SEQ ID NO:689, SEQ ID NO:755, SEQ ID NO:816, SEQ ID NO:876, SEQ ID NO:927, SEQ ID NO:978, SEQ ID NO:1013, SEQ ID NO:1029, SEQ ID NO:1055, SEQ ID NO:1060, SEQ ID NO:1093, SEQ ID NO:1128, SEQ ID NO:1157, SEQ ID NO:1191, SEQ ID NO:1212,
1300 SEQ ID NO:1240, SEQ ID NO:1258, SEQ ID NO:1274, SEQ ID NO:1288, SEQ ID NO:1302, SEQ ID NO:1309, SEQ ID NO:1321, SEQ ID NO:1329, SEQ ID NO:1338, SEQ ID NO:1348, SEQ ID NO:1359, SEQ ID NO:1366, SEQ ID NO:1374, SEQ ID NO:1380, SEQ ID NO:1386, SEQ ID NO:1394, SEQ ID NO:1401, SEQ ID
1305 NO:1408, SEQ ID NO:1411, SEQ ID NO:1418, SEQ ID NO:1426, SEQ ID NO:1436, SEQ ID NO:1443, SEQ ID NO:1450, SEQ ID NO:1457, SEQ ID NO:1460, SEQ ID NO:1467, SEQ ID NO:1471, SEQ ID NO:1473, SEQ ID NO:1478, SEQ ID NO:1487, SEQ ID NO:1489, SEQ ID NO:1494, SEQ ID NO:1499, SEQ ID NO:
1310 1501, SEQ ID NO:1506, SEQ ID NO:1508, SEQ ID NO:1510, SEQ ID NO:1511, SEQ ID NO:1516, SEQ ID NO:1520, SEQ ID NO:1526, SEQ ID NO:1532, SEQ ID NO:1537, SEQ ID NO:1540, SEQ ID NO:1545, SEQ ID NO:1547, SEQ ID NO:1549, SEQ ID NO:1551, SEQ ID NO:1553, SEQ ID NO:1555, SEQ ID NO:1558,

1315 SEQ ID NO:1563, SEQ ID NO:1566, SEQ ID NO:1569, SEQ ID
 NO:1571, SEQ ID NO:1576, SEQ ID NO:1580,SEQ ID NO:1584,
 SEQ ID NO:1587, SEQ ID NO:1591, SEQ ID NO:1594, SEQID
 NO:1596, SEQ ID NO:1599, SEQ ID NO:1601, SEQ ID NO:1603,
 SEQ ID NO:1604, SEQ ID NO:1605, SEQ ID NO:1607, SEQ ID
 1320 NO:1612, SEQ ID NO:1615, SEQ ID NO:1617, SEQ ID NO:1619,
 SEQ ID NO:1622, SEQ ID NO:1624,SEQ ID NO:1626, SEQ ID
 NO:1627, SEQ ID NO:1629, SEQ ID NO:1632, SEQID NO:1635,
 SEQ ID NO:1636, SEQ ID NO:1637, SEQ ID NO:1639, SEQ
 IDNO:1640, SEQ ID NO:1643, SEQ ID NO:1646, SEQ ID NO:
 1325 1649, SEQ ID NO:1652, SEQ ID NO:1655, SEQ ID NO:1658,
 SEQ ID NO:1660, SEQ ID NO:1662, SEQ ID NO:1664, SEQ ID
 NO:1666, SEQ ID NO:1668, SEQ ID NO:1669, SEQ ID NO:1670,
 SEQ ID NO:1672, SEQ ID NO:1673, SEQ ID NO:1675, SEQ
 IDNO:1677, SEQ ID NO:1680, SEQ ID NO:1682, SEQ ID NO:
 1330 1683, SEQ ID NO:1685, SEQ ID NO:1688, SEQ ID NO:1690,
 SEQ ID NO:1691, SEQ ID NO:1694, SEQ ID NO:1696, SEQ ID
 NO:1699, SEQ ID NO:1700, SEQ ID NO:1701,SEQ ID NO:1704,
 SEQ ID NO:1705, SEQ ID NO:1706, SEQ ID NO:1707, SEQID
 NO:1708, SEQ ID NO:1709, SEQ ID NO:1710, SEQ ID NO:1711,
 1335 SEQ ID NO:1712, SEQ ID NO:1713, SEQ ID NO:1715, SEQ ID
 NO:1716, SEQ ID NO:1717, SEQ ID NO:1718,, SEQ ID NO:
 1719, SEQ ID NO:1720, SEQ ID NO:1721, SEQ ID NO:1722,
 SEQ ID NO:1723, SEQ ID NO:1724, SEQ ID NO:1725, SEQ ID
 NO:1726, SEQ ID NO:1727, SEQ ID NO:1728, SEQ ID NO:1729,
 1340 SEQ IDNO:1730, SEQ ID NO:1731, SEQ ID NO:1732, SEQ ID
 NO:1733, SEQ ID NO:1734, SEQ ID NO:1735, SEQ ID NO:1736,
 SEQ ID NO:1737, SEQ ID NO:1738, SEQ ID NO:1739, SEQ ID
 NO:1740, SEQ ID NO:1741, SEQ ID NO:1742,SEQ ID NO:1743,
 SEQ ID NO:1744, SEQ ID NO:1745, SEQ ID NO:1746, SEQID
 1345 NO:1747, SEQ ID NO:1748, SEQ ID NO:1749, SEQ ID NO:1750,
 SEQ ID NO:1751, SEQ ID NO:1752, SEQ ID NO:1753, SEQ ID
 NO:1754, SEQ ID NO:1755, SEQ ID NO:1756, SEQ ID NO:1757,
 SEQ ID NO:1758, SEQ ID NO:1759,SEQ ID NO:1760, SEQ ID

NO:1761, SEQ ID NO:1762, SEQ ID NO:1763, SEQ ID NO:1764,
1350 SEQ ID NO:1765, SEQ ID NO:1766, SEQ ID NO:1767, SEQ
IDNO:1768, SEQ ID NO:1769, SEQ ID NO:1770, SEQ ID NO:
1771, SEQ ID NO:1772, SEQ ID NO:1773, SEQ ID NO:1774,
SEQ ID NO:1775, SEQ ID NO:1776, SEQ ID NO:1777, SEQ ID
NO:1778, SEQ ID NO:1779, SEQ ID NO:1780, SEQ ID NO:1781,
1355 SEQ ID NO:1782, SEQ ID NO:1783, SEQ ID NO:1784, SEQ
IDNO:1785, SEQ ID NO:1786, SEQ ID NO:1787, SEQ ID NO:
1788, SEQ ID NO:1789, SEQ ID NO:1790, SEQ ID NO:1791,
SEQ ID NO:1792, SEQ ID NO:1793, SEQ ID NO:1794, SEQ ID
NO:1795, SEQ ID NO:1796, SEQ ID NO:1797,SEQ ID NO:1798,
1360 SEQ ID NO:1799, SEQ ID NO:1800, SEQ ID NO:1801, SEQ ID
NO:1802, SEQ ID NO:1803, SEQ ID NO:1804, SEQ ID NO:1805,
SEQ ID NO:1806, SEQ ID NO:1807, SEQ ID NO:1808, SEQ ID
NO:1809, SEQ ID NO:1810, SEQ ID NO:1811, SEQ ID NO:1812,
SEQ ID NO:1813, SEQ ID NO:1814,SEQ ID NO:1815, SEQ ID
1365 NO:1816, SEQ ID NO:1817, SEQ ID NO:1818, SEQ ID NO:1819,
SEQ ID NO:1820, SEQ ID NO:1821, SEQ ID NO:1822, SEQ
IDNO:1823, SEQ ID NO:1824, SEQ ID NO:1825, SEQ ID NO:
1826, SEQ ID NO:1827, SEQ ID NO:1828, SEQ ID NO:1829,
SEQ ID NO:1830, SEQ ID NO:1831, SEQ ID NO:1832, SEQ ID
1370 NO:1833, SEQ ID NO:1834, SEQ ID NO:1835, SEQ ID NO:1836,
SEQ ID NO:1837, SEQ ID NO:1838, SEQ ID NO:1839, SEQ
IDNO:1840, SEQ ID NO:1841, SEQ ID NO:1842, SEQ ID NO:
1843, SEQ ID NO:1844, SEQ ID NO:1845, SEQ ID NO:1846,
SEQ ID NO:1847, SEQ ID NO:1848, SEQ ID NO:1849, SEQ ID
1375 NO:1850, SEQ ID NO:1851, SEQ ID NO:1852,SEQ ID NO:1853,
SEQ ID NO:1854, SEQ ID NO:1855, SEQ ID NO:1856, SEQ ID
NO:1857, SEQ ID NO:1858, SEQ ID NO:1859, SEQ ID NO:1860,
SEQ ID NO:1861, SEQ ID NO:1862, SEQ ID NO:1863, SEQ ID
NO:1864, SEQ ID NO:1865, and SEQ ID NO:1866

1380 (b) a partial sequence in the nucleotide sequences set
forth in (a);

(c) a complementary nucleotide sequence to the

nucleotide sequence set forth in (a) or (b); or

(d) a nucleotide sequence hybridizing to the nucleotide
1385 sequences set forth in (a), (b) or (c) under a stringent condition.

These nucleic-acid molecules of the present invention
include a large number of O-157 specific genes, [wherein] the
genes encode proteins or polypeptides specific to O-157.
[0012]

1390 Accordingly, the present invention relates to a
nucleic-acid molecule which is a nucleic-acid molecule encoding
a polypeptide specific to enterohemorrhagic pathogenic E. coli
O-157:H7 and encodes

(a) an amino acid sequence selected from a group
1395 comprising the following SEQ IDs or a fragment thereof, SEQ
ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID
NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:
10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID
NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID
1400 NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID
NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID
NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID
NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID
NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID
1405 NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID
NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID
NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID
NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID
NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID
1410 NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID
NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID
NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID
NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID
NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID
1415 NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID
NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID

NO:86, SEQ ID NO:87, SEQ IDNO:88, SEQ ID NO:89, SEQ ID
 NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQID NO:93, SEQ ID
 NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97,SEQ ID
 1420 NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ
 ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105,
 SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:
 109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID
 NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116,
 1425 SEQ ID NO:117, SEQ ID NO:118, SEQ ID NO:119, SEQ ID NO:
 120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID
 NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127,
 SEQ ID NO:128, SEQ ID NO:129, SEQ ID NO:130, SEQ ID NO:
 131, SEQ ID NO:133, SEQ ID NO:134, SEQ ID NO:135, SEQ ID
 1430 NO:136, SEQ ID NO:137, SEQ ID NO:138, SEQ ID NO:139,
 SEQ ID NO:140, SEQ ID NO:141, SEQ ID NO:142, SEQ ID NO:
 143, SEQ ID NO:144, SEQ ID NO:145, SEQ ID NO:146, SEQ ID
 NO:147, SEQ ID NO:148, SEQ ID NO:149, SEQ ID NO:150,
 SEQ ID NO:151, SEQ ID NO:152, SEQ ID NO:153, SEQ ID NO:
 1435 154, SEQ ID NO:155, SEQ ID NO:156, SEQ ID NO:157, SEQ ID
 NO:158, SEQ ID NO:159, SEQ ID NO:160, SEQ ID NO:161,
 SEQ ID NO:162, SEQ ID NO:163, SEQ ID NO:164, SEQ ID NO:
 165, SEQ ID NO:166, SEQ ID NO:167, SEQ ID NO:168, SEQ ID
 NO:169, SEQ ID NO:170, SEQ ID NO:171, SEQ ID NO:172,
 1440 SEQ ID NO:173, SEQ ID NO:174, SEQ ID NO:175, SEQ ID NO:
 176, SEQ ID NO:177, SEQ ID NO:178, SEQ ID NO:179, SEQ ID
 NO:180, SEQ ID NO:181, SEQ ID NO:182, SEQ ID NO:183,
 SEQ ID NO:184, SEQ ID NO:185, SEQ ID NO:186, SEQ ID NO:
 187, SEQ ID NO:188, SEQ ID NO:189, SEQ ID NO:190, SEQ ID
 1445 NO:191, SEQ ID NO:192, SEQ ID NO:193, SEQ ID NO:194,
 SEQ ID NO:195, SEQ ID NO:196, SEQ ID NO:197, SEQ ID NO:
 198, SEQ ID NO:199, SEQ ID NO:200, SEQ ID NO:201, SEQ ID
 NO:202, SEQ ID NO:203, SEQ ID NO:204, SEQ ID NO:205,
 SEQ ID NO:206, SEQ ID NO:207, SEQ ID NO:208, SEQ ID NO:
 1450 209, SEQ ID NO:210, SEQ ID NO:211, SEQ ID NO:212, SEQ ID

NO:213, SEQ ID NO:214, SEQ ID NO:215, SEQ ID NO:216,
SEQ ID NO:217, SEQ ID NO:218, SEQ ID NO:219, SEQ ID NO:
220, SEQ ID NO:221, SEQ ID NO:222, SEQ ID NO:223, SEQ ID
1455 NO:224, SEQ ID NO:225, SEQ ID NO:226, SEQ ID NO:227,
SEQ ID NO:228, SEQ ID NO:229, SEQ ID NO:230, SEQ ID NO:
231, SEQ ID NO:232, SEQ ID NO:233, SEQ ID NO:234, SEQ ID
NO:235, SEQ ID NO:236, SEQ ID NO:237, SEQ ID NO:238,
SEQ ID NO:239, SEQ ID NO:240, SEQ ID NO:241, SEQ ID NO:
242, SEQ ID NO:243, SEQ ID NO:245, SEQ ID NO:246, SEQ ID
1460 NO:247, SEQ ID NO:248, SEQ ID NO:249, SEQ ID NO:250,
SEQ ID NO:251, SEQ ID NO:252, SEQ ID NO:253, SEQ ID NO:
254, SEQ ID NO:255, SEQ ID NO:256, SEQ ID NO:257, SEQ ID
NO:258, SEQ ID NO:259, SEQ ID NO:260, SEQ ID NO:261,
SEQ ID NO:262, SEQ ID NO:263, SEQ ID NO:264, SEQ ID NO:
1465 265, SEQ ID NO:266, SEQ ID NO:267, SEQ ID NO:268, SEQ ID
NO:269, SEQ ID NO:270, SEQ ID NO:271, SEQ ID NO:272,
SEQ ID NO:273, SEQ ID NO:274, SEQ ID NO:275, SEQ ID NO:
276, SEQ ID NO:277, SEQ ID NO:278, SEQ ID NO:279, SEQ ID
NO:280, SEQ ID NO:281, SEQ ID NO:282, SEQ ID NO:283,
1470 SEQ ID NO:284, SEQ ID NO:285, SEQ ID NO:286, SEQ ID NO:
287, SEQ ID NO:288, SEQ ID NO:289, SEQ ID NO:290, SEQ ID
NO:291, SEQ ID NO:292, SEQ ID NO:293, SEQ ID NO:294,
SEQ ID NO:295, SEQ ID NO:296, SEQ ID NO:297, SEQ ID NO:
298, SEQ ID NO:299, SEQ ID NO:300, SEQ ID NO:301, SEQ ID
1475 NO:302, SEQ ID NO:303, SEQ ID NO:304, SEQ ID NO:305,
SEQ ID NO:306, SEQ ID NO:307, SEQ ID NO:308, SEQ ID NO:
309, SEQ ID NO:310, SEQ ID NO:311, SEQ ID NO:312, SEQ ID
NO:313, SEQ ID NO:314, SEQ ID NO:315, SEQ ID NO:316,
SEQ ID NO:317, SEQ ID NO:318, SEQ ID NO:319, SEQ ID NO:
1480 320, SEQ ID NO:321, SEQ ID NO:322, SEQ ID NO:323, SEQ ID
NO:324, SEQ ID NO:325, SEQ ID NO:326, SEQ ID NO:327,
SEQ ID NO:328, SEQ ID NO:329, SEQ ID NO:330, SEQ ID NO:
331, SEQ ID NO:332, SEQ ID NO:333, SEQ ID NO:334, SEQ ID
NO:335, SEQ ID NO:336, SEQ ID NO:338, SEQ ID NO:339,

1485 SEQ ID NO:340, SEQ ID NO:341, SEQ ID NO:342, SEQ ID NO:
343, SEQ ID NO:344, SEQ ID NO:345, SEQ ID NO:346, SEQ ID
NO:347, SEQ ID NO:348, SEQ ID NO:349, SEQ ID NO:350,
SEQ ID NO:351, SEQ ID NO:352, SEQ ID NO:353, SEQ ID NO:
354, SEQ ID NO:355, SEQ ID NO:356, SEQ ID NO:357, SEQ ID
1490 NO:358, SEQ ID NO:359, SEQ ID NO:360, SEQ ID NO:361,
SEQ ID NO:362, SEQ ID NO:363, SEQ ID NO:364, SEQ ID NO:
365, SEQ ID NO:366, SEQ ID NO:367, SEQ ID NO:368, SEQ ID
NO:369, SEQ ID NO:370, SEQ ID NO:371, SEQ ID NO:372,
SEQ ID NO:373, SEQ ID NO:374, SEQ ID NO:375, SEQ ID NO:
1495 376, SEQ ID NO:377, SEQ ID NO:378, SEQ ID NO:379, SEQ ID
NO:380, SEQ ID NO:381, SEQ ID NO:382, SEQ ID NO:383,
SEQ ID NO:384, SEQ ID NO:385, SEQ ID NO:386, SEQ ID NO:
387, SEQ ID NO:388, SEQ ID NO:389, SEQ ID NO:390, SEQ ID
NO:391, SEQ ID NO:392, SEQ ID NO:393, SEQ ID NO:394,
1500 SEQ ID NO:395, SEQ ID NO:396, SEQ ID NO:397, SEQ ID NO:
398, SEQ ID NO:399, SEQ ID NO:400, SEQ ID NO:401, SEQ ID
NO:402, SEQ ID NO:403, SEQ ID NO:404, SEQ ID NO:405,
SEQ ID NO:406, SEQ ID NO:407, SEQ ID NO:408, SEQ ID NO:
409, SEQ ID NO:411, SEQ ID NO:412, SEQ ID NO:413, SEQ ID
1505 NO:414, SEQ ID NO:415, SEQ ID NO:416, SEQ ID NO:417,
SEQ ID NO:418, SEQ ID NO:419, SEQ ID NO:420, SEQ ID NO:
421, SEQ ID NO:422, SEQ ID NO:423, SEQ ID NO:424, SEQ ID
NO:425, SEQ ID NO:426, SEQ ID NO:427, SEQ ID NO:428,
SEQ ID NO:429, SEQ ID NO:430, SEQ ID NO:431, SEQ ID NO:
1510 432, SEQ ID NO:433, SEQ ID NO:434, SEQ ID NO:435, SEQ
ID NO:436, SEQ ID NO:437, SEQ ID NO:438, SEQ ID NO:439,
SEQ ID NO:440, SEQ ID NO:441, SEQ ID NO:442, SEQ ID NO:
443, SEQ ID NO:444, SEQ ID NO:445, SEQ ID NO:446, SEQ ID
NO:447, SEQ ID NO:448, SEQ ID NO:449, SEQ ID NO:450,
1515 SEQ ID NO:451, SEQ ID NO:452, SEQ ID NO:453, SEQ ID NO:
454, SEQ ID NO:455, SEQ ID NO:456, SEQ ID NO:457, SEQ ID
NO:458, SEQ ID NO:459, SEQ ID NO:460, SEQ ID NO:461,
SEQ ID NO:462, SEQ ID NO:463, SEQ ID NO:464, SEQ ID NO:

465, SEQ ID NO:466, SEQ ID NO:467, SEQ ID NO:468, SEQ ID
1520 NO:469, SEQ ID NO:470, SEQ ID NO:471, SEQ ID NO:472,
SEQ ID NO:473, SEQ ID NO:474, SEQ ID NO:475, SEQ ID NO:
476, SEQ ID NO:477, SEQ ID NO:478, SEQ ID NO:479, SEQ ID
NO:480, SEQ ID NO:481, SEQ ID NO:482, SEQ ID NO:483,
SEQ ID NO:485, SEQ ID NO:486, SEQ ID NO:487, SEQ ID NO:
1525 488, SEQ ID NO:489, SEQ ID NO:490, SEQ ID NO:491, SEQ ID
NO:492, SEQ ID NO:493, SEQ ID NO:494, SEQ ID NO:495,
SEQ ID NO:496, SEQ ID NO:497, SEQ ID NO:498, SEQ ID NO:
499, SEQ ID NO:500, SEQ ID NO:501, SEQ ID NO:502, SEQ ID
NO:503, SEQ ID NO:504, SEQ ID NO:505, SEQ ID NO:506,
1530 SEQ ID NO:507, SEQ ID NO:508, SEQ ID NO:509, SEQ ID NO:
510, SEQ ID NO:511, SEQ ID NO:512, SEQ ID NO:513, SEQ ID
NO:514, SEQ ID NO:515, SEQ ID NO:516, SEQ ID NO:517,
SEQ ID NO:518, SEQ ID NO:519, SEQ ID NO:520, SEQ ID NO:
521, SEQ ID NO:522, SEQ ID NO:523, SEQ ID NO:524, SEQ ID
1535 NO:525, SEQ ID NO:526, SEQ ID NO:527, SEQ ID NO:528,
SEQ ID NO:529, SEQ ID NO:530, SEQ ID NO:531, SEQ ID NO:
532, SEQ ID NO:533, SEQ ID NO:534, SEQ ID NO:535, SEQ ID
NO:536, SEQ ID NO:537, SEQ ID NO:538, SEQ ID NO:539,
SEQ ID NO:540, SEQ ID NO:541, SEQ ID NO:542, SEQ ID NO:
1540 543, SEQ ID NO:544, SEQ ID NO:545, SEQ ID NO:546, SEQ ID
NO:547, SEQ ID NO:548, SEQ ID NO:549, SEQ ID NO:550,
SEQ ID NO:551, SEQ ID NO:552, SEQ ID NO:553, SEQ ID NO:
555, SEQ ID NO:556, SEQ ID NO:557, SEQ ID NO:558, SEQ ID
NO:559, SEQ ID NO:560, SEQ ID NO:561, SEQ ID NO:562,
1545 SEQ ID NO:563, SEQ ID NO:564, SEQ ID NO:565, SEQ ID NO:
566, SEQ ID NO:567, SEQ ID NO:568, SEQ ID NO:569, SEQ ID
NO:570, SEQ ID NO:571, SEQ ID NO:572, SEQ ID NO:573,
SEQ ID NO:574, SEQ ID NO:575, SEQ ID NO:576, SEQ ID NO:
577, SEQ ID NO:578, SEQ ID NO:579, SEQ ID NO:580, SEQ ID
1550 NO:581, SEQ ID NO:582, SEQ ID NO:583, SEQ ID NO:584,
SEQ ID NO:585, SEQ ID NO:586, SEQ ID NO:587, SEQ ID NO:
588, SEQ ID NO:589, SEQ ID NO:590, SEQ ID NO:591, SEQ ID

NO:592, SEQ ID NO:593, SEQ ID NO:594, SEQ ID NO:595,
SEQ ID NO:596, SEQ ID NO:597, SEQ ID NO:598, SEQ ID NO:
1555 599, SEQ ID NO:600, SEQ ID NO:601, SEQ ID NO:602, SEQ ID
NO:603, SEQ ID NO:604, SEQ ID NO:605, SEQ ID NO:606,
SEQ ID NO:607, SEQ ID NO:608, SEQ ID NO:609, SEQ ID NO:
610, SEQ ID NO:611, SEQ ID NO:612, SEQ ID NO:613, SEQ ID
NO:614, SEQ ID NO:615, SEQ ID NO:616, SEQ ID NO:617,
1560 SEQ ID NO:618, SEQ ID NO:619, SEQ ID NO:620, SEQ ID NO:
621, SEQ ID NO:622, SEQ ID NO:623, SEQ ID NO:624, SEQ ID
NO:625, SEQ ID NO:626, SEQ ID NO:627, SEQ ID NO:628,
SEQ ID NO:629, SEQ ID NO:631, SEQ ID NO:632, SEQ ID NO:
633, SEQ ID NO:634, SEQ ID NO:635, SEQ ID NO:636, SEQ ID
1565 NO:637, SEQ ID NO:638, SEQ ID NO:639, SEQ ID NO:640,
SEQ ID NO:641, SEQ ID NO:642, SEQ ID NO:643, SEQ ID NO:
644, SEQ ID NO:645, SEQ ID NO:646, SEQ ID NO:647, SEQ ID
NO:648, SEQ ID NO:649, SEQ ID NO:650, SEQ ID NO:651,
SEQ ID NO:652, SEQ ID NO:653, SEQ ID NO:654, SEQ ID NO:
1570 655, SEQ ID NO:656, SEQ ID NO:657, SEQ ID NO:658, SEQ ID
NO:659, SEQ ID NO:660, SEQ ID NO:661, SEQ ID NO:662,
SEQ ID NO:663, SEQ ID NO:664, SEQ ID NO:665, SEQ ID NO:
666, SEQ ID NO:667, SEQ ID NO:668, SEQ ID NO:669, SEQ ID
NO:670, SEQ ID NO:671, SEQ ID NO:672, SEQ ID NO:673,
1575 SEQ ID NO:674, SEQ ID NO:675, SEQ ID NO:676, SEQ ID NO:
677, SEQ ID NO:678, SEQ ID NO:679, SEQ ID NO:680, SEQ ID
NO:681, SEQ ID NO:682, SEQ ID NO:683, SEQ ID NO:684,
SEQ ID NO:685, SEQ ID NO:686, SEQ ID NO:687, SEQ ID NO:
688, SEQ ID NO:690, SEQ ID NO:691, SEQ ID NO:692, SEQ ID
1580 NO:693, SEQ ID NO:694, SEQ ID NO:695, SEQ ID NO:696,
SEQ ID NO:697, SEQ ID NO:698, SEQ ID NO:699, SEQ ID N
O:700, SEQ ID NO:701, SEQ ID NO:702, SEQ ID NO:703, SEQ
ID NO:704, SEQ ID NO:705, SEQ ID NO:706, SEQ ID NO:707,
SEQ ID NO:708, SEQ ID NO:709, SEQ ID NO:710, SEQ ID NO:
1585 711, SEQ ID NO:712, SEQ ID NO:713, SEQ ID NO:714, SEQ ID
NO:715, SEQ ID NO:716, SEQ ID NO:717, SEQ ID NO:718,

SEQ ID NO:719, SEQ ID NO:720, SEQ ID NO:721, SEQ ID NO:
722, SEQ ID NO:723, SEQ ID NO:724, SEQ ID NO:725, SEQ ID
NO:726, SEQ ID NO:727, SEQ ID NO:728, SEQ ID NO:729,
1590 SEQ ID NO:730, SEQ ID NO:731, SEQ ID NO:732, SEQ ID NO:
733, SEQ ID NO:734, SEQ ID NO:735, SEQ ID NO:736, SEQ ID
NO:737, SEQ ID NO:738, SEQ ID NO:739, SEQ ID NO:740,
SEQ ID NO:741, SEQ ID NO:742, SEQ ID NO:743, SEQ ID NO:
744, SEQ ID NO:745, SEQ ID NO:746, SEQ ID NO:747, SEQ ID
1595 NO:748, SEQ ID NO:749, SEQ ID NO:750, SEQ ID NO:751,
SEQ ID NO:752, SEQ ID NO:753, SEQ ID NO:754, SEQ ID NO:
756, SEQ ID NO:757, SEQ ID NO:758, SEQ ID NO:759, SEQ ID
NO:760, SEQ ID NO:761, SEQ ID NO:762, SEQ ID NO:763,
SEQ ID NO:764, SEQ ID NO:765, SEQ ID NO:766, SEQ ID NO:
1600 767, SEQ ID NO:768, SEQ ID NO:769, SEQ ID NO:770, SEQ ID
NO:771, SEQ ID NO:772, SEQ ID NO:773, SEQ ID NO:774,
SEQ ID NO:775, SEQ ID NO:776, SEQ ID NO:777, SEQ ID NO:
778, SEQ ID NO:779, SEQ ID NO:780, SEQ ID NO:781, SEQ ID
NO:782, SEQ ID NO:783, SEQ ID NO:784, SEQ ID NO:785,
1605 SEQ ID NO:786, SEQ ID NO:787, SEQ ID NO:788, SEQ ID NO:
789, SEQ ID NO:790, SEQ ID NO:791, SEQ ID NO:792, SEQ ID
NO:793, SEQ ID NO:794, SEQ ID NO:795, SEQ ID NO:796,
SEQ ID NO:797, SEQ ID NO:798, SEQ ID NO:799, SEQ ID NO:
800, SEQ ID NO:801, SEQ ID NO:802, SEQ ID NO:803, SEQ ID
1610 NO:804, SEQ ID NO:805, SEQ ID NO:806, SEQ ID NO:807,
SEQ ID NO:808, SEQ ID NO:809, SEQ ID NO:810, SEQ ID NO:
811, SEQ ID NO:812, SEQ ID NO:813, SEQ ID NO:814, SEQ ID
NO:815, SEQ ID NO:817, SEQ ID NO:818, SEQ ID NO:819,
SEQ ID NO:820, SEQ ID NO:821, SEQ ID NO:822, SEQ ID NO:
1615 823, SEQ ID NO:824, SEQ ID NO:825, SEQ ID NO:826, SEQ ID
NO:827, SEQ ID NO:828, SEQ ID NO:829, SEQ ID NO:830,
SEQ ID NO:831, SEQ ID NO:832, SEQ ID NO:833, SEQ ID NO:
834, SEQ ID NO:835, SEQ ID NO:836, SEQ ID NO:837, SEQ ID
NO:838, SEQ ID NO:839, SEQ ID NO:840, SEQ ID NO:841,
1620 SEQ ID NO:842, SEQ ID NO:843, SEQ ID NO:844, SEQ ID NO:

845, SEQ ID NO:846, SEQ ID NO:847, SEQ ID NO:848, SEQ ID
 NO:849, SEQ ID NO:850, SEQ ID NO:851, SEQ ID NO:852,
 SEQ ID NO:853, SEQ ID NO:854, SEQ ID NO:855, SEQ ID NO:
 856, SEQ ID NO:857, SEQ ID NO:858, SEQ ID NO:859, SEQ ID
 1625 NO:860, SEQ ID NO:861, SEQ ID NO:862, SEQ ID NO:863,
 SEQ ID NO:864, SEQ ID NO:865, SEQ ID NO:866, SEQ ID NO:
 867, SEQ ID NO:868, SEQ ID NO:869, SEQ ID NO:870, SEQ ID
 NO:871, SEQ ID NO:872, SEQ ID NO:873, SEQ ID NO:874,
 SEQ ID NO:875, SEQ ID NO:877, SEQ ID NO:878, SEQ ID NO:
 1630 879, SEQ ID NO:880, SEQ ID NO:881, SEQ ID NO:882, SEQ ID
 NO:883, SEQ ID NO:884, SEQ ID NO:885, SEQ ID NO:886,
 SEQ ID NO:887, SEQ ID NO:888, SEQ ID NO:889, SEQ ID NO:
 890, SEQ ID NO:891, SEQ ID NO:892, SEQ ID NO:893, SEQ ID
 NO:894, SEQ ID NO:895, SEQ ID NO:896, SEQ ID NO:897,
 1635 SEQ ID NO:898, SEQ ID NO:899, SEQ ID NO:900, SEQ ID NO:
 901, SEQ ID NO:902, SEQ ID NO:903, SEQ ID NO:904, SEQ ID
 NO:905, SEQ ID NO:906, SEQ ID NO:907, SEQ ID NO:908,
 SEQ ID NO:909, SEQ ID NO:910, SEQ ID NO:911, SEQ ID NO:
 912, SEQ ID NO:913, SEQ ID NO:914, SEQ ID NO:915, SEQ ID
 1640 NO:916, SEQ ID NO:917, SEQ ID NO:918, SEQ ID NO:919,
 SEQ ID NO:920, SEQ ID NO:921, SEQ ID NO:922, SEQ ID NO:
 923, SEQ ID NO:924, SEQ ID NO:925, SEQ ID NO:926, SEQ ID
 NO:928, SEQ ID NO:929, SEQ ID NO:930, SEQ ID NO:931,
 SEQ ID NO:932, SEQ ID NO:933, SEQ ID NO:934, SEQ ID NO:
 1645 935, SEQ ID NO:936, SEQ ID NO:937, SEQ ID NO:938, SEQ ID
 NO:939, SEQ ID NO:940, SEQ ID NO:941, SEQ ID NO:942,
 SEQ ID NO:943, SEQ ID NO:944, SEQ ID NO:945, SEQ ID NO:
 946, SEQ ID NO:947, SEQ ID NO:948, SEQ ID NO:949, SEQ ID
 NO:950, SEQ ID NO:951, SEQ ID NO:952, SEQ ID NO:953,
 1650 SEQ ID NO:954, SEQ ID NO:955, SEQ ID NO:956, SEQ ID NO:
 957, SEQ ID NO:958, SEQ ID NO:959, SEQ ID NO:960, SEQ ID
 NO:961, SEQ ID NO:962, SEQ ID NO:963, SEQ ID NO:964, SE
 Q ID NO:965, SEQ ID NO:966, SEQ ID NO:967, SEQ ID NO:
 968, SEQ ID NO:969, SEQ ID NO:970, SEQ ID NO:971, SEQ ID

1655 NO:972, SEQ ID NO:973, SEQ ID NO:974, SEQ ID NO:975,
 SEQ ID NO:976, SEQ ID NO:977, SEQ ID NO:979, SEQ ID NO:
 980, SEQ ID NO:981, SEQ ID NO:982, SEQ ID NO:983, SEQ ID
 NO:984, SEQ ID NO:985, SEQ ID NO:986, SEQ ID NO:987,
 SEQ ID NO:988, SEQ ID NO:989, SEQ ID NO:990, SEQ ID NO:
 1660 991, SEQ ID NO:992, SEQ ID NO:993, SEQ ID NO:994, SEQ ID
 NO:995, SEQ ID NO:996, SEQ ID NO:997, SEQ ID NO:998,
 SEQ ID NO:999, SEQ ID NO:1000, SEQ ID NO:1001,SEQ ID
 NO:1002, SEQ ID NO:1003, SEQ ID NO:1004, SEQ ID NO:1005,
 SEQID NO:1006, SEQ ID NO:1007, SEQ ID NO:1008, SEQ ID
 1665 NO:1009, SEQ IDNO:1010, SEQ ID NO:1011, SEQ ID NO:1012,
 SEQ ID NO:1014, SEQ ID NO:1015, SEQ ID NO:1016, SEQ ID
 NO:1017, SEQ ID NO:1018, SEQ ID NO:1019, SEQ ID NO:1020,
 SEQ ID NO:1021, SEQ ID NO:1022, SEQ ID NO:1023, SEQ ID
 NO:1024, SEQ ID NO:1025, SEQ ID NO:1026, SEQ ID NO:1027,
 1670 SEQ IDNO:1028, SEQ ID NO:1030, SEQ ID NO:1031, SEQ ID
 NO:1032, SEQ ID NO:1033, SEQ ID NO:1034, SEQ ID NO:1035,
 SEQ ID NO:1036, SEQ ID NO:1037, SEQ ID NO:1038, SEQ ID
 NO:1039, SEQ ID NO:1040, SEQ ID NO:1041,SEQ ID NO:1042,
 SEQ ID NO:1043, SEQ ID NO:1044, SEQ ID NO:1045, SEQID
 1675 NO:1046, SEQ ID NO:1047, SEQ ID NO:1048, SEQ ID NO:1049,
 SEQ ID NO:1050, SEQ ID NO:1051, SEQ ID NO:1052, SEQ ID
 NO:1053, SEQ ID NO:1054, SEQ ID NO:1056, SEQ ID NO:1057,
 SEQ ID NO:1058, SEQ ID NO:1059,SEQ ID NO:1061, SEQ ID
 NO:1062, SEQ ID NO:1063, SEQ ID NO:1064, SEQID NO:1065,
 1680 SEQ ID NO:1066, SEQ ID NO:1067, SEQ ID NO:1068, SEQ
 IDNO:1069, SEQ ID NO:1070, SEQ ID NO:1071, SEQ ID NO:
 1072, SEQ ID NO:1073, SEQ ID NO:1074, SEQ ID NO:1075,
 SEQ ID NO:1076, SEQ ID NO:1077, SEQ ID NO:1078, SEQ ID
 NO:1079, SEQ ID NO:1080, SEQ ID NO:1081, SEQ ID NO:1082,
 1685 SEQ ID NO:1083, SEQ ID NO:1084, SEQ ID NO:1085, SEQ
 IDNO:1086, SEQ ID NO:1087, SEQ ID NO:1088, SEQ ID NO:
 1089, SEQ ID NO:1090, SEQ ID NO:1091, SEQ ID NO:1092,
 SEQ ID NO:1094, SEQ ID NO:1095, SEQ ID NO:1096, SEQ ID

NO:1097, SEQ ID NO:1098, SEQ ID NO:1099,SEQ ID NO:1100,
 1690 SEQ ID NO:1101, SEQ ID NO:1102, SEQ ID NO:1103, SEQID
 NO:1104, SEQ ID NO:1105, SEQ ID NO:1106, SEQ ID NO:1107,
 SEQ ID NO:1108, SEQ ID NO:1109, SEQ ID NO:1110, SEQ ID
 NO:1111, SEQ ID NO:1112, SEQ ID NO:1113, SEQ ID NO:1114,
 SEQ ID NO:1115, SEQ ID NO:1116,SEQ ID NO:1117, SEQ ID
 1695 NO:1118, SEQ ID NO:1119, SEQ ID NO:1120, SEQID NO:1121,
 SEQ ID NO:1122, SEQ ID NO:1123, SEQ ID NO:1124, SEQ
 IDNO:1125, SEQ ID NO:1126, SEQ ID NO:1127, SEQ ID NO:
 1129, SEQ ID NO:1130, SEQ ID NO:1131, SEQ ID NO:1132,
 SEQ ID NO:1133, SEQ ID NO:1134, SEQ ID NO:1135, SEQ ID
 1700 NO:1136, SEQ ID NO:1137, SEQ ID NO:1138, SEQ ID NO:1139,
 SEQ ID NO:1140, SEQ ID NO:1141, SEQ ID NO:1142, SEQ
 IDNO:1143, SEQ ID NO:1144, SEQ ID NO:1145, SEQ ID NO:
 1146, SEQ ID NO:1147, SEQ ID NO:1148, SEQ ID NO:1149,
 SEQ ID NO:1150, SEQ ID NO:1151, SEQ ID NO:1152, SEQ ID
 1705 NO:1153, SEQ ID NO:1154, SEQ ID NO:1155,SEQ ID NO:1156,
 SEQ ID NO:1158, SEQ ID NO:1159, SEQ ID NO:1160, SEQID
 NO:1161, SEQ ID NO:1162, SEQ ID NO:1163, SEQ ID NO:1164,
 SEQ ID NO:1165, SEQ ID NO:1166, SEQ ID NO:1167, SEQ ID
 NO:1168, SEQ ID NO:1169, SEQ ID NO:1170, SEQ ID NO:1171,
 1710 SEQ ID NO:1172, SEQ ID NO:1173,SEQ ID NO:1174, SEQ ID
 NO:1175, SEQ ID NO:1176, SEQ ID NO:1177, SEQID NO:1178,
 SEQ ID NO:1179, SEQ ID NO:1180, SEQ ID NO:1181, SEQ
 IDNO:1182, SEQ ID NO:1183, SEQ ID NO:1184, SEQ ID NO:
 1185, SEQ ID NO:1186, SEQ ID NO:1187, SEQ ID NO:1188,
 1715 SEQ ID NO:1189, SEQ ID NO:1190, SEQ ID NO:1192, SEQ ID
 NO:1193, SEQ ID NO:1194, SEQ ID NO:1195, SEQ ID NO:1196,
 SEQ ID NO:1197, SEQ ID NO:1198, SEQ ID NO:1199, SEQ
 IDNO:1200, SEQ ID NO:1201, SEQ ID NO:1202, SEQ ID NO:
 1203, SEQ ID NO:1204, SEQ ID NO:1205, SEQ ID NO:1206,
 1720 SEQ ID NO:1207, SEQ ID NO:1208, SEQ ID NO:1209, SEQ ID
 NO:1210, SEQ ID NO:1211, SEQ ID NO:1213,SEQ ID NO:1214,
 SEQ ID NO:1215, SEQ ID NO:1216, SEQ ID NO:1217, SEQID

NO:1218, SEQ ID NO:1219, SEQ ID NO:1220, SEQ ID NO:1221,
 SEQ ID NO:1222, SEQ ID NO:1223, SEQ ID NO:1224, SEQ ID
 1725 NO:1225, SEQ ID NO:1226, SEQ ID NO:1227, SEQ ID NO:1228,
 SEQ ID NO:1229, SEQ ID NO:1230, SEQ ID NO:1231, SEQ ID
 NO:1232, SEQ ID NO:1233, SEQ ID NO:1234, SEQ ID NO:1235,
 SEQ ID NO:1236, SEQ ID NO:1237, SEQ ID NO:1238, SEQ
 IDNO:1239, SEQ ID NO:1241, SEQ ID NO:1242, SEQ ID NO:
 1730 1243, SEQ ID NO:1244, SEQ ID NO:1245, SEQ ID NO:1246,
 SEQ ID NO:1247, SEQ ID NO:1248, SEQ ID NO:1249, SEQ ID
 NO:1250, SEQ ID NO:1251, SEQ ID NO:1252, SEQ ID NO:1253,
 SEQ ID NO:1254, SEQ ID NO:1255, SEQ ID NO:1256, SEQ
 IDNO:1257, SEQ ID NO:1259, SEQ ID NO:1260, SEQ ID NO:
 1735 1261, SEQ ID NO:1262, SEQ ID NO:1263, SEQ ID NO:1264,
 SEQ ID NO:1265, SEQ ID NO:1266, SEQ ID NO:1267, SEQ ID
 NO:1268, SEQ ID NO:1269, SEQ ID NO:1270, SEQ ID NO:1271,
 SEQ ID NO:1272, SEQ ID NO:1273, SEQ ID NO:1275, SEQ ID
 NO:1276, SEQ ID NO:1277, SEQ ID NO:1278, SEQ ID NO:1279,
 1740 SEQ ID NO:1280, SEQ ID NO:1281, SEQ ID NO:1282, SEQ ID
 NO:1283, SEQ ID NO:1284, SEQ ID NO:1285, SEQ ID NO:1286,
 SEQ ID NO:1287, SEQ ID NO:1289, SEQ ID NO:1290, SEQ ID
 NO:1291, SEQ ID NO:1292, SEQ ID NO:1293, SEQ ID NO:1294,
 SEQ ID NO:1295, SEQ ID NO:1296, SEQ ID NO:1297, SEQ
 1745 IDNO:1298, SEQ ID NO:1299, SEQ ID NO:1300, SEQ ID NO:
 1301, SEQ ID NO:1303, SEQ ID NO:1304, SEQ ID NO:1305,
 SEQ ID NO:1306, SEQ ID NO:1307, SEQ ID NO:1308, SEQ ID
 NO:1310, SEQ ID NO:1311, SEQ ID NO:1312, SEQ ID NO:1313,
 SEQ ID NO:1314, SEQ ID NO:1315, SEQ ID NO:1316, SEQ
 1750 IDNO:1317, SEQ ID NO:1318, SEQ ID NO:1319, SEQ ID NO:
 1320, SEQ ID NO:1322, SEQ ID NO:1323, SEQ ID NO:1324,
 SEQ ID NO:1325, SEQ ID NO:1326, SEQ ID NO:1327, SEQ ID
 NO:1328, SEQ ID NO:1330, SEQ ID NO:1331, SEQ ID NO:1332,
 SEQ ID NO:1333, SEQ ID NO:1334, SEQ ID NO:1335, SEQ ID
 1755 NO:1336, SEQ ID NO:1337, SEQ ID NO:1339, SEQ ID NO:1340,
 SEQ ID NO:1341, SEQ ID NO:1342, SEQ ID NO:1343, SEQ ID

NO:1344, SEQ ID NO:1345, SEQ ID NO:1346, SEQ ID NO:1347,
SEQ ID NO:1349, SEQ ID NO:1350,SEQ ID NO:1351, SEQ ID
NO:1352, SEQ ID NO:1353, SEQ ID NO:1354, SEQ ID NO:1355,
1760 SEQ ID NO:1356, SEQ ID NO:1357, SEQ ID NO:1358, SEQ
IDNO:1360, SEQ ID NO:1361, SEQ ID NO:1362, SEQ ID NO:
1363, SEQ ID NO:1364, SEQ ID NO:1365, SEQ ID NO:1367,
SEQ ID NO:1368, SEQ ID NO:1369, SEQ ID NO:1370, SEQ ID
NO:1371, SEQ ID NO:1375, SEQ ID NO:1376, SEQ ID NO:1377,
1765 SEQ ID NO:1378, SEQ ID NO:1379, SEQ ID NO:1381, SEQ
IDNO:1382, SEQ ID NO:1383, SEQ ID NO:1384, SEQ ID NO:
1385, SEQ ID NO:1387, SEQ ID NO:1388, SEQ ID NO:1389,
SEQ ID NO:1390, SEQ ID NO:1391, SEQ ID NO:1392, SEQ ID
NO:1393, SEQ ID NO:1395, SEQ ID NO:1396,SEQ ID NO:1397,
1770 SEQ ID NO:1398, SEQ ID NO:1399, SEQ ID NO:1400, SEQ ID
NO:1402, SEQ ID NO:1403, SEQ ID NO:1404, SEQ ID NO:1405,
SEQ ID NO:1406, SEQ ID NO:1407, SEQ ID NO:1409, SEQ ID
NO:1410, SEQ ID NO:1412, SEQ ID NO:1413, SEQ ID NO:1414,
SEQ ID NO:1415, SEQ ID NO:1416,SEQ ID NO:1417, SEQ ID
1775 NO:1419, SEQ ID NO:1420, SEQ ID NO:1421, SEQ ID NO:1422,
SEQ ID NO:1423, SEQ ID NO:1424, SEQ ID NO:1425, SEQ
IDNO:1427, SEQ ID NO:1428, SEQ ID NO:1429, SEQ ID NO:
1430, SEQ ID NO:1431, SEQ ID NO:1432, SEQ ID NO:1433,
SEQ ID NO:1434, SEQ ID NO:1435, SEQ ID NO:1437, SEQ ID
1780 NO:1438, SEQ ID NO:1439, SEQ ID NO:1440, SEQ ID NO:1441,
SEQ ID NO:1442, SEQ ID NO:1444, SEQ ID NO:1445, SEQ
IDNO:1446, SEQ ID NO:1447, SEQ ID NO:1448, SEQ ID NO:
1449, SEQ ID NO:1451, SEQ ID NO:1452, SEQ ID NO:1453,
SEQ ID NO:1454, SEQ ID NO:1455, SEQ ID NO:1456, SEQ ID
1785 NO:1458, SEQ ID NO:1459, SEQ ID NO:1461,SEQ ID NO:1462,
SEQ ID NO:1463, SEQ ID NO:1464, SEQ ID NO:1465, SEQ ID
NO:1466, SEQ ID NO:1468, SEQ ID NO:1469, SEQ ID NO:1470,
SEQ ID NO:1472, SEQ ID NO:1474, SEQ ID NO:1475, SEQ ID
NO:1476, SEQ ID NO:1477, SEQ ID NO:1479, SEQ ID NO:1480,
1790 SEQ ID NO:1481, SEQ ID NO:1482,SEQ ID NO:1483, SEQ ID

NO:1484, SEQ ID NO:1485, SEQ ID NO:1486, SEQID NO:1488,
 SEQ ID NO:1490, SEQ ID NO:1491, SEQ ID NO:1492, SEQ
 IDNO:1493, SEQ ID NO:1495, SEQ ID NO:1496, SEQ ID NO:
 1497, SEQ ID NO:1498, SEQ ID NO:1500, SEQ ID NO:1502,
 1795 SEQ ID NO:1503, SEQ ID NO:1504, SEQ ID NO:1505, SEQ
 ID NO:1507, SEQ ID NO:1509, SEQ ID NO:1512, SEQ ID NO:
 1513, SEQ ID NO:1514, SEQ ID NO:1515, SEQ ID NO:1517,
 SEQ IDNO:1518, SEQ ID NO:1519, SEQ ID NO:1521, SEQ ID
 NO:1522, SEQ ID NO:1523, SEQ ID NO:1524, SEQ ID NO:1525,
 1800 SEQ ID NO:1527, SEQ ID NO:1528, SEQ ID NO:1529, SEQ ID
 NO:1530, SEQ ID NO:1531, SEQ ID NO:1533,SEQ ID NO:1534,
 SEQ ID NO:1535, SEQ ID NO:1536, SEQ ID NO:1538, SEQID
 NO:1539, SEQ ID NO:1541, SEQ ID NO:1542, SEQ ID NO:1543,
 SEQ ID NO:1544, SEQ ID NO:1546, SEQ ID NO:1548, SEQ ID
 1805 NO:1550, SEQ ID NO:1552, SEQ ID NO:1554, SEQ ID NO:1556,
 SEQ ID NO:1557, SEQ ID NO:1559,SEQ ID NO:1560, SEQ ID
 NO:1561, SEQ ID NO:1562, SEQ ID NO:1564, SEQID NO:1565,
 SEQ ID NO:1567, SEQ ID NO:1568, SEQ ID NO:1570, SEQ
 IDNO:1572, SEQ ID NO:1573, SEQ ID NO:1574, SEQ ID NO:
 1810 1575, SEQ ID NO:1577, SEQ ID NO:1578, SEQ ID NO:1579,
 SEQ ID NO:1581, SEQ ID NO:1582, SEQ ID NO:1583, SEQ ID
 NO:1585, SEQ ID NO:1586, SEQ ID NO:1588, SEQ ID NO:1589,
 SEQ ID NO:1590, SEQ ID NO:1592, SEQ ID NO:1593, SEQ
 IDNO:1595, SEQ ID NO:1597, SEQ ID NO:1598, SEQ ID NO:
 1815 1600, SEQ ID NO:1602, SEQ ID NO:1606, SEQ ID NO:1608,
 SEQ ID NO:1609, SEQ ID NO:1610, SEQ ID NO:1611, SEQ ID
 NO:1613, SEQ ID NO:1614, SEQ ID NO:1616,SEQ ID NO:1618,
 SEQ ID NO:1620, SEQ ID NO:1621, SEQ ID NO:1623, SEQID
 NO:1625, SEQ ID NO:1628, SEQ ID NO:1630, SEQ ID NO:1631,
 1820 SEQ ID NO:1633, SEQ ID NO:1634, SEQ ID NO:1638, SEQ ID
 NO:1641, SEQ ID NO:1642, SEQ ID NO:1644, SEQ ID NO:1645,
 SEQ ID NO:1647, SEQ ID NO:1648,SEQ ID NO:1650, SEQ ID
 NO:1651, SEQ ID NO:1653, SEQ ID NO:1654, SEQID NO:1656,
 SEQ ID NO:1657, SEQ ID NO:1659, SEQ ID NO:1661, SEQ

1825 IDNO:1663, SEQ ID NO:1665, SEQ ID NO:1667, SEQ ID NO:
1671, SEQ ID NO:1674, SEQ ID NO:1676, SEQ ID NO:1678,
SEQ ID NO:1679, SEQ ID NO:1681, SEQ ID NO:1684, SEQ ID
NO:1686, SEQ ID NO:1687, SEQ ID NO:1689, SEQ ID NO:1692,
SEQ ID NO:1693, SEQ ID NO:1695, SEQ ID NO:1697, SEQ
1830 IDNO:1698, SEQ ID NO:1702, and SEQ ID NO:1703

, or (b) a polypeptide comprising an amino acid sequence
in the nucleotide sequences set forth in (a) in which several
amino acids are deleted, replaced or added.
[0013]

1835 In another embodiment, the present invention relates to a
polypeptide specific to enterohemorrhagic pathogenic E. coli
O-157:H7.

In a preferred embodiment, the present invention relates to a
polypeptide specific to O-157:H7 comprising

1840 (a) an amino acid sequence selected from a group
comprising the following SEQ IDs or a fragment thereof: SEQ
ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID
NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:
10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID
1845 NO:14, SEQ ID NO:15, SEQ IDNO:16, SEQ ID NO:17, SEQ ID
NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQID NO:21, SEQ ID
NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25,SEQ ID
NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID
NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID
1850 NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID
NO:38, SEQ ID NO:39, SEQ IDNO:40, SEQ ID NO:41, SEQ ID
NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQID NO:45, SEQ ID
NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49,SEQ ID
NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID
1855 NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID
NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID
NO:62, SEQ ID NO:63, SEQ IDNO:64, SEQ ID NO:65, SEQ ID
NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQID NO:69, SEQ ID

NO: 70, SEQ ID NO: 71, SEQ ID NO: 72, SEQ ID NO: 73, SEQ ID
 1860 NO: 74, SEQ ID NO: 75, SEQ ID NO: 76, SEQ ID NO: 77, SEQ ID
 NO: 78, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 81, SEQ ID
 NO: 82, SEQ ID NO: 83, SEQ ID NO: 84, SEQ ID NO: 85, SEQ ID
 NO: 86, SEQ ID NO: 87, SEQ ID NO: 88, SEQ ID NO: 89, SEQ ID
 1865 NO: 90, SEQ ID NO: 91, SEQ ID NO: 92, SEQ ID NO: 93, SEQ ID
 NO: 94, SEQ ID NO: 95, SEQ ID NO: 96, SEQ ID NO: 97, SEQ ID
 NO: 98, SEQ ID NO: 99, SEQ ID NO: 100, SEQ ID NO: 101, SEQ
 ID NO: 102, SEQ ID NO: 103, SEQ ID NO: 104, SEQ ID NO: 105,
 SEQ ID NO: 106, SEQ ID NO: 107, SEQ ID NO: 108, SEQ ID NO:
 109, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112, SEQ ID
 1870 NO: 113, SEQ ID NO: 114, SEQ ID NO: 115, SEQ ID NO: 116,
 SEQ ID NO: 117, SEQ ID NO: 118, SEQ ID NO: 119, SEQ ID NO:
 120, SEQ ID NO: 121, SEQ ID NO: 122, SEQ ID NO: 123, SEQ ID
 NO: 124, SEQ ID NO: 125, SEQ ID NO: 126, SEQ ID NO: 127,
 SEQ ID NO: 128, SEQ ID NO: 129, SEQ ID NO: 130, SEQ ID NO:
 1875 131, SEQ ID NO: 133, SEQ ID NO: 134, SEQ ID NO: 135, SEQ ID
 NO: 136, SEQ ID NO: 137, SEQ ID NO: 138, SEQ ID NO: 139,
 SEQ ID NO: 140, SEQ ID NO: 141, SEQ ID NO: 142, SEQ ID NO:
 143, SEQ ID NO: 144, SEQ ID NO: 145, SEQ ID NO: 146, SEQ ID
 NO: 147, SEQ ID NO: 148, SEQ ID NO: 149, SEQ ID NO: 150,
 1880 SEQ ID NO: 151, SEQ ID NO: 152, SEQ ID NO: 153, SEQ ID NO:
 154, SEQ ID NO: 155, SEQ ID NO: 156, SEQ ID NO: 157, SEQ ID
 NO: 158, SEQ ID NO: 159, SEQ ID NO: 160, SEQ ID NO: 161,
 SEQ ID NO: 162, SEQ ID NO: 163, SEQ ID NO: 164, SEQ ID NO:
 165, SEQ ID NO: 166, SEQ ID NO: 167, SEQ ID NO: 168, SEQ ID
 1885 NO: 169, SEQ ID NO: 170, SEQ ID NO: 171, SEQ ID NO: 172,
 SEQ ID NO: 173, SEQ ID NO: 174, SEQ ID NO: 175, SEQ ID NO:
 176, SEQ ID NO: 177, SEQ ID NO: 178, SEQ ID NO: 179, SEQ ID
 NO: 180, SEQ ID NO: 181, SEQ ID NO: 182, SEQ ID NO: 183,
 SEQ ID NO: 184, SEQ ID NO: 185, SEQ ID NO: 186, SEQ ID NO:
 1890 187, SEQ ID NO: 188, SEQ ID NO: 189, SEQ ID NO: 190, SEQ ID
 NO: 191, SEQ ID NO: 192, SEQ ID NO: 193, SEQ ID NO: 194,
 SEQ ID NO: 195, SEQ ID NO: 196, SEQ ID NO: 197, SEQ ID NO:

198, SEQ ID NO:199, SEQ ID NO:200, SEQ ID NO:201, SEQ ID
 NO:202, SEQ ID NO:203, SEQ ID NO:204, SEQ ID NO:205,
 1895 SEQ ID NO:206, SEQ ID NO:207, SEQ ID NO:208, SEQ ID NO:
 209, SEQ ID NO:210, SEQ ID NO:211, SEQ ID NO:212, SEQ ID
 NO:213, SEQ ID NO:214, SEQ ID NO:215, SEQ ID NO:216,
 SEQ ID NO:217, SEQ ID NO:218, SEQ ID NO:219, SEQ ID NO:
 220, SEQ ID NO:221, SEQ ID NO:222, SEQ ID NO:223, SEQ ID
 1900 NO:224, SEQ ID NO:225, SEQ ID NO:226, SEQ ID NO:227,
 SEQ ID NO:228, SEQ ID NO:229, SEQ ID NO:230, SEQ ID NO:
 231, SEQ ID NO:232, SEQ ID NO:233, SEQ ID NO:234, SEQ
 ID NO:235, SEQ ID NO:236, SEQ ID NO:237, SEQ ID NO:238,
 SEQ ID NO:239, SEQ ID NO:240, SEQ ID NO:241, SEQ ID NO:
 1905 242, SEQ ID NO:243, SEQ ID NO:245, SEQ ID NO:246, SEQ ID
 NO:247, SEQ ID NO:248, SEQ ID NO:249, SEQ ID NO:250,
 SEQ ID NO:251, SEQ ID NO:252, SEQ ID NO:253, SEQ ID NO:
 254, SEQ ID NO:255, SEQ ID NO:256, SEQ ID NO:257, SEQ ID
 NO:258, SEQ ID NO:259, SEQ ID NO:260, SEQ ID NO:261,
 1910 SEQ ID NO:262, SEQ ID NO:263, SEQ ID NO:264, SEQ ID NO:
 265, SEQ ID NO:266, SEQ ID NO:267, SEQ ID NO:268, SEQ ID
 NO:269, SEQ ID NO:270, SEQ ID NO:271, SEQ ID NO:272,
 SEQ ID NO:273, SEQ ID NO:274, SEQ ID NO:275, SEQ ID NO:
 276, SEQ ID NO:277, SEQ ID NO:278, SEQ ID NO:279, SEQ ID
 1915 NO:280, SEQ ID NO:281, SEQ ID NO:282, SEQ ID NO:283,
 SEQ ID NO:284, SEQ ID NO:285, SEQ ID NO:286, SEQ ID NO:
 287, SEQ ID NO:288, SEQ ID NO:289, SEQ ID NO:290, SEQ ID
 NO:291, SEQ ID NO:292, SEQ ID NO:293, SEQ ID NO:294,
 SEQ ID NO:295, SEQ ID NO:296, SEQ ID NO:297, SEQ ID NO:
 1920 298, SEQ ID NO:299, SEQ ID NO:300, SEQ ID NO:301, SEQ ID
 NO:302, SEQ ID NO:303, SEQ ID NO:304, SEQ ID NO:305,
 SEQ ID NO:306, SEQ ID NO:307, SEQ ID NO:308, SEQ ID NO:
 309, SEQ ID NO:310, SEQ ID NO:311, SEQ ID NO:312, SEQ ID
 NO:313, SEQ ID NO:314, SEQ ID NO:315, SEQ ID NO:316,
 1925 SEQ ID NO:317, SEQ ID NO:318, SEQ ID NO:319, SEQ ID NO:
 320, SEQ ID NO:321, SEQ ID NO:322, SEQ ID NO:323, SEQ ID

NO:324, SEQ ID NO:325, SEQ ID NO:326, SEQ ID NO:327,
 SEQ ID NO:328, SEQ ID NO:329, SEQ ID NO:330, SEQ ID NO:
 331, SEQ ID NO:332, SEQ ID NO:333, SEQ ID NO:334, SEQ ID
 1930 NO:335, SEQ ID NO:336, SEQ ID NO:338, SEQ ID NO:339,
 SEQ ID NO:340, SEQ ID NO:341, SEQ ID NO:342, SEQ ID NO:
 343, SEQ ID NO:344, SEQ ID NO:345, SEQ ID NO:346, SEQ ID
 NO:347, SEQ ID NO:348, SEQ ID NO:349, SEQ ID NO:350,
 SEQ ID NO:351, SEQ ID NO:352, SEQ ID NO:353, SEQ ID NO:
 1935 354, SEQ ID NO:355, SEQ ID NO:356, SEQ ID NO:357, SEQ ID
 NO:358, SEQ ID NO:359, SEQ ID NO:360, SEQ ID NO:361,
 SEQ ID NO:362, SEQ ID NO:363, SEQ ID NO:364, SEQ ID NO:
 365, SEQ ID NO:366, SEQ ID NO:367, SEQ ID NO:368, SEQ ID
 NO:369, SEQ ID NO:370, SEQ ID NO:371, SEQ ID NO:372,
 1940 SEQ ID NO:373, SEQ ID NO:374, SEQ ID NO:375, SEQ ID NO:
 376, SEQ ID NO:377, SEQ ID NO:378, SEQ ID NO:379, SEQ ID
 NO:380, SEQ ID NO:381, SEQ ID NO:382, SEQ ID NO:383,
 SEQ ID NO:384, SEQ ID NO:385, SEQ ID NO:386, SEQ ID NO:
 387, SEQ ID NO:388, SEQ ID NO:389, SEQ ID NO:390, SEQ ID
 1945 NO:391, SEQ ID NO:392, SEQ ID NO:393, SEQ ID NO:394,
 SEQ ID NO:395, SEQ ID NO:396, SEQ ID NO:397, SEQ ID NO:
 398, SEQ ID NO:399, SEQ ID NO:400, SEQ ID NO:401, SEQ ID
 NO:402, SEQ ID NO:403, SEQ ID NO:404, SEQ ID NO:405,
 SEQ ID NO:406, SEQ ID NO:407, SEQ ID NO:408, SEQ ID NO:
 1950 409, SEQ ID NO:411, SEQ ID NO:412, SEQ ID NO:413, SEQ ID
 NO:414, SEQ ID NO:415, SEQ ID NO:416, SEQ ID NO:417,
 SEQ ID NO:418, SEQ ID NO:419, SEQ ID NO:420, SEQ ID NO:
 421, SEQ ID NO:422, SEQ ID NO:423, SEQ ID NO:424, SEQ ID
 NO:425, SEQ ID NO:426, SEQ ID NO:427, SEQ ID NO:428,
 1955 SEQ ID NO:429, SEQ ID NO:430, SEQ ID NO:431, SEQ ID NO:
 432, SEQ ID NO:433, SEQ ID NO:434, SEQ ID NO:435, SEQ ID
 NO:436, SEQ ID NO:437, SEQ ID NO:438, SEQ ID NO:439,
 SEQ ID NO:440, SEQ ID NO:441, SEQ ID NO:442, SEQ ID NO:
 443, SEQ ID NO:444, SEQ ID NO:445, SEQ ID NO:446, SEQ ID
 1960 NO:447, SEQ ID NO:448, SEQ ID NO:449, SEQ ID NO:450,

SEQ ID NO:451, SEQ ID NO:452, SEQ ID NO:453, SEQ ID NO:
 454, SEQ ID NO:455, SEQ ID NO:456, SEQ ID NO:457, SEQ ID
 NO:458, SEQ ID NO:459, SEQ ID NO:460, SEQ ID NO:461,
 1965 SEQ ID NO:462, SEQ ID NO:463, SEQ ID NO:464, SEQ ID NO:
 465, SEQ ID NO:466, SEQ ID NO:467, SEQ ID NO:468, SEQ ID
 NO:469, SEQ ID NO:470, SEQ ID NO:471, SEQ ID NO:472,
 SEQ ID NO:473, SEQ ID NO:474, SEQ ID NO:475, SEQ ID NO:
 476, SEQ ID NO:477, SEQ ID NO:478, SEQ ID NO:479, SEQ ID
 NO:480, SEQ ID NO:481, SEQ ID NO:482, SEQ ID NO:483,
 1970 SEQ ID NO:485, SEQ ID NO:486, SEQ ID NO:487, SEQ ID NO:
 488, SEQ ID NO:489, SEQ ID NO:490, SEQ ID NO:491, SEQ ID
 NO:492, SEQ ID NO:493, SEQ ID NO:494, SEQ ID NO:495,
 SEQ ID NO:496, SEQ ID NO:497, SEQ ID NO:498, SEQ ID NO:
 499, SEQ ID NO:500, SEQ ID NO:501, SEQ ID NO:502, SEQ ID
 1975 NO:503, SEQ ID NO:504, SEQ ID NO:505, SEQ ID NO:506,
 SEQ ID NO:507, SEQ ID NO:508, SEQ ID NO:509, SEQ ID NO:
 510, SEQ ID NO:511, SEQ ID NO:512, SEQ ID NO:513, SEQ ID
 NO:514, SEQ ID NO:515, SEQ ID NO:516, SEQ ID NO:517,
 SEQ ID NO:518, SEQ ID NO:519, SEQ ID NO:520, SEQ ID NO:
 1980 521, SEQ ID NO:522, SEQ ID NO:523, SEQ ID NO:524, SEQ ID
 NO:525, SEQ ID NO:526, SEQ ID NO:527, SEQ ID NO:528,
 SEQ ID NO:529, SEQ ID NO:530, SEQ ID NO:531, SEQ ID NO:
 532, SEQ ID NO:533, SEQ ID NO:534, SEQ ID NO:535, SEQ ID
 NO:536, SEQ ID NO:537, SEQ ID NO:538, SEQ ID NO:539,
 1985 SEQ ID NO:540, SEQ ID NO:541, SEQ ID NO:542, SEQ ID NO:
 543, SEQ ID NO:544, SEQ ID NO:545, SEQ ID NO:546, SEQ ID
 NO:547, SEQ ID NO:548, SEQ ID NO:549, SEQ ID NO:550,
 SEQ ID NO:551, SEQ ID NO:552, SEQ ID NO:553, SEQ ID NO:
 555, SEQ ID NO:556, SEQ ID NO:557, SEQ ID NO:558, SEQ ID
 1990 NO:559, SEQ ID NO:560, SEQ ID NO:561, SEQ ID NO:562,
 SEQ ID NO:563, SEQ ID NO:564, SEQ ID NO:565, SEQ ID NO:
 566, SEQ ID NO:567, SEQ ID NO:568, SEQ ID NO:569, SEQ ID
 NO:570, SEQ ID NO:571, SEQ ID NO:572, SEQ ID NO:573,
 SEQ ID NO:574, SEQ ID NO:575, SEQ ID NO:576, SEQ ID NO:

1995 577, SEQ ID NO:578, SEQ ID NO:579, SEQ ID NO:580, SEQ ID
NO:581, SEQ ID NO:582, SEQ ID NO:583, SEQ ID NO:584,
SEQ ID NO:585, SEQ ID NO:586, SEQ ID NO:587, SEQ ID NO:
588, SEQ ID NO:589, SEQ ID NO:590, SEQ ID NO:591, SEQ ID
NO:592, SEQ ID NO:593, SEQ ID NO:594, SEQ ID NO:595,
2000 SEQ ID NO:596, SEQ ID NO:597, SEQ ID NO:598, SEQ ID NO:
599, SEQ ID NO:600, SEQ ID NO:601, SEQ ID NO:602, SEQ ID
NO:603, SEQ ID NO:604, SEQ ID NO:605, SEQ ID NO:606,
SEQ ID NO:607, SEQ ID NO:608, SEQ ID NO:609, SEQ ID NO:
610, SEQ ID NO:611, SEQ ID NO:612, SEQ ID NO:613, SEQ ID
2005 NO:614, SEQ ID NO:615, SEQ ID NO:616, SEQ ID NO:617,
SEQ ID NO:618, SEQ ID NO:619, SEQ ID NO:620, SEQ ID NO:
621, SEQ ID NO:622, SEQ ID NO:623, SEQ ID NO:624, SEQ ID
NO:625, SEQ ID NO:626, SEQ ID NO:627, SEQ ID NO:628,
SEQ ID NO:629, SEQ ID NO:631, SEQ ID NO:632, SEQ ID NO:
2010 633, SEQ ID NO:634, SEQ ID NO:635, SEQ ID NO:636, SEQ ID
NO:637, SEQ ID NO:638, SEQ ID NO:639, SEQ ID NO:640,
SEQ ID NO:641, SEQ ID NO:642, SEQ ID NO:643, SEQ ID NO:
644, SEQ ID NO:645, SEQ ID NO:646, SEQ ID NO:647, SEQ ID
NO:648, SEQ ID NO:649, SEQ ID NO:650, SEQ ID NO:651,
2015 SEQ ID NO:652, SEQ ID NO:653, SEQ ID NO:654, SEQ ID NO:
655, SEQ ID NO:656, SEQ ID NO:657, SEQ ID NO:658, SEQ ID
NO:659, SEQ ID NO:660, SEQ ID NO:661, SEQ ID NO:662,
SEQ ID NO:663, SEQ ID NO:664, SEQ ID NO:665, SEQ ID NO:
666, SEQ ID NO:667, SEQ ID NO:668, SEQ ID NO:669, SEQ ID
2020 NO:670, SEQ ID NO:671, SEQ ID NO:672, SEQ ID NO:673,
SEQ ID NO:674, SEQ ID NO:675, SEQ ID NO:676, SEQ ID NO:
677, SEQ ID NO:678, SEQ ID NO:679, SEQ ID NO:680, SEQ ID
NO:681, SEQ ID NO:682, SEQ ID NO:683, SEQ ID NO:684,
SEQ ID NO:685, SEQ ID NO:686, SEQ ID NO:687, SEQ ID NO:
2025 688, SEQ ID NO:690, SEQ ID NO:691, SEQ ID NO:692, SEQ ID
NO:693, SEQ ID NO:694, SEQ ID NO:695, SEQ ID NO:696,
SEQ ID NO:697, SEQ ID NO:698, SEQ ID NO:699, SEQ ID NO:
700, SEQ ID NO:701, SEQ ID NO:702, SEQ ID NO:703, SEQ ID

NO:704, SEQ ID NO:705, SEQ ID NO:706, SEQ ID NO:707,
2030 SEQ ID NO:708, SEQ ID NO:709, SEQ ID NO:710, SEQ ID NO:
711, SEQ ID NO:712, SEQ ID NO:713, SEQ ID NO:714, SEQ ID
NO:715, SEQ ID NO:716, SEQ ID NO:717, SEQ ID NO:718,
SEQ ID NO:719, SEQ ID NO:720, SEQ ID NO:721, SEQ ID NO:
722, SEQ ID NO:723, SEQ ID NO:724, SEQ ID NO:725, SEQ ID
2035 NO:726, SEQ ID NO:727, SEQ ID NO:728, SEQ ID NO:729,
SEQ ID NO:730, SEQ ID NO:731, SEQ ID NO:732, SEQ ID NO:
733, SEQ ID NO:734, SEQ ID NO:735, SEQ ID NO:736, SEQ ID
NO:737, SEQ ID NO:738, SEQ ID NO:739, SEQ ID NO:740,
SEQ ID NO:741, SEQ ID NO:742, SEQ ID NO:743, SEQ ID NO:
2040 744, SEQ ID NO:745, SEQ ID NO:746, SEQ ID NO:747, SEQ ID
NO:748, SEQ ID NO:749, SEQ ID NO:750, SEQ ID NO:751,
SEQ ID NO:752, SEQ ID NO:753, SEQ ID NO:754, SEQ ID NO:
756, SEQ ID NO:757, SEQ ID NO:758, SEQ ID NO:759, SEQ ID
NO:760, SEQ ID NO:761, SEQ ID NO:762, SEQ ID NO:763, SE
2045 Q ID NO:764, SEQ ID NO:765, SEQ ID NO:766, SEQ ID NO:
767, SEQ ID NO:768, SEQ ID NO:769, SEQ ID NO:770, SEQ ID
NO:771, SEQ ID NO:772, SEQ ID NO:773, SEQ ID NO:774,
SEQ ID NO:775, SEQ ID NO:776, SEQ ID NO:777, SEQ ID NO:
778, SEQ ID NO:779, SEQ ID NO:780, SEQ ID NO:781, SEQ ID
2050 NO:782, SEQ ID NO:783, SEQ ID NO:784, SEQ ID NO:785,
SEQ ID NO:786, SEQ ID NO:787, SEQ ID NO:788, SEQ ID NO:
789, SEQ ID NO:790, SEQ ID NO:791, SEQ ID NO:792, SEQ ID
NO:793, SEQ ID NO:794, SEQ ID NO:795, SEQ ID NO:796,
SEQ ID NO:797, SEQ ID NO:798, SEQ ID NO:799, SEQ ID NO:
2055 800, SEQ ID NO:801, SEQ ID NO:802, SEQ ID NO:803, SEQ ID
NO:804, SEQ ID NO:805, SEQ ID NO:806, SEQ ID NO:807,
SEQ ID NO:808, SEQ ID NO:809, SEQ ID NO:810, SEQ ID NO:
811, SEQ ID NO:812, SEQ ID NO:813, SEQ ID NO:814, SEQ ID
NO:815, SEQ ID NO:817, SEQ ID NO:818, SEQ ID NO:819,
2060 SEQ ID NO:820, SEQ ID NO:821, SEQ ID NO:822, SEQ ID NO:
823, SEQ ID NO:824, SEQ ID NO:825, SEQ ID NO:826, SEQ ID
NO:827, SEQ ID NO:828, SEQ ID NO:829, SEQ ID NO:830,

SEQ ID NO:831, SEQ ID NO:832, SEQ ID NO:833, SEQ ID NO:
834, SEQ ID NO:835, SEQ ID NO:836, SEQ ID NO:837, SEQ ID
2065 NO:838, SEQ ID NO:839, SEQ ID NO:840, SEQ ID NO:841,
SEQ ID NO:842, SEQ ID NO:843, SEQ ID NO:844, SEQ ID NO:
845, SEQ ID NO:846, SEQ ID NO:847, SEQ ID NO:848, SEQ ID
NO:849, SEQ ID NO:850, SEQ ID NO:851, SEQ ID NO:852,
SEQ ID NO:853, SEQ ID NO:854, SEQ ID NO:855, SEQ ID NO:
2070 856, SEQ ID NO:857, SEQ ID NO:858, SEQ ID NO:859, SEQ ID
NO:860, SEQ ID NO:861, SEQ ID NO:862, SEQ ID NO:863,
SEQ ID NO:864, SEQ ID NO:865, SEQ ID NO:866, SEQ ID NO:
867, SEQ ID NO:868, SEQ ID NO:869, SEQ ID NO:870, SEQ ID
NO:871, SEQ ID NO:872, SEQ ID NO:873, SEQ ID NO:874,
2075 SEQ ID NO:875, SEQ ID NO:877, SEQ ID NO:878, SEQ ID NO:
879, SEQ ID NO:880, SEQ ID NO:881, SEQ ID NO:882, SEQ ID
NO:883, SEQ ID NO:884, SEQ ID NO:885, SEQ ID NO:886,
SEQ ID NO:887, SEQ ID NO:888, SEQ ID NO:889, SEQ ID NO:
890, SEQ ID NO:891, SEQ ID NO:892, SEQ ID NO:893, SEQ ID
2080 NO:894, SEQ ID NO:895, SEQ ID NO:896, SEQ ID NO:897,
SEQ ID NO:898, SEQ ID NO:899, SEQ ID NO:900, SEQ ID NO:
901, SEQ ID NO:902, SEQ ID NO:903, SEQ ID NO:904, SEQ ID
NO:905, SEQ ID NO:906, SEQ ID NO:907, SEQ ID NO:908,
SEQ ID NO:909, SEQ ID NO:910, SEQ ID NO:911, SEQ ID NO:
2085 912, SEQ ID NO:913, SEQ ID NO:914, SEQ ID NO:915, SEQ ID
NO:916, SEQ ID NO:917, SEQ ID NO:918, SEQ ID NO:919,
SEQ ID NO:920, SEQ ID NO:921, SEQ ID NO:922, SEQ ID NO:
923, SEQ ID NO:924, SEQ ID NO:925, SEQ ID NO:926, SEQ ID
NO:928, SEQ ID NO:929, SEQ ID NO:930, SEQ ID NO:931,
2090 SEQ ID NO:932, SEQ ID NO:933, SEQ ID NO:934, SEQ ID NO:
935, SEQ ID NO:936, SEQ ID NO:937, SEQ ID NO:938, SEQ ID
NO:939, SEQ ID NO:940, SEQ ID NO:941, SEQ ID NO:942,
SEQ ID NO:943, SEQ ID NO:944, SEQ ID NO:945, SEQ ID NO:
946, SEQ ID NO:947, SEQ ID NO:948, SEQ ID NO:949, SEQ ID
2095 NO:950, SEQ ID NO:951, SEQ ID NO:952, SEQ ID NO:953,
SEQ ID NO:954, SEQ ID NO:955, SEQ ID NO:956, SEQ ID NO:

957, SEQ ID NO:958, SEQ ID NO:959, SEQ ID NO:960, SEQ ID
 NO:961, SEQ ID NO:962, SEQ ID NO:963, SEQ ID NO:964,
 SEQ ID NO:965, SEQ ID NO:966, SEQ ID NO:967, SEQ ID NO:
 2100 968, SEQ ID NO:969, SEQ ID NO:970, SEQ ID NO:971, SEQ ID
 NO:972, SEQ ID NO:973, SEQ ID NO:974, SEQ ID NO:975,
 SEQ ID NO:976, SEQ ID NO:977, SEQ ID NO:979, SEQ ID NO:
 980, SEQ ID NO:981, SEQ ID NO:982, SEQ ID NO:983, SEQ ID
 NO:984, SEQ ID NO:985, SEQ ID NO:986, SEQ ID NO:987,
 2105 SEQ ID NO:988, SEQ ID NO:989, SEQ ID NO:990, SEQ ID NO:
 991, SEQ ID NO:992, SEQ ID NO:993, SEQ ID NO:994, SEQ ID
 NO:995, SEQ ID NO:996, SEQ ID NO:997, SEQ ID NO:998,
 SEQ ID NO:999, SEQ ID NO:1000, SEQ ID NO:1001,SEQ ID
 NO:1002, SEQ ID NO:1003, SEQ ID NO:1004, SEQ ID NO:1005,
 2110 SEQID NO:1006, SEQ ID NO:1007, SEQ ID NO:1008, SEQ ID
 NO:1009, SEQ IDNO:1010, SEQ ID NO:1011, SEQ ID NO:1012,
 SEQ ID NO:1014, SEQ ID NO:1015, SEQ ID NO:1016, SEQ ID
 NO:1017, SEQ ID NO:1018, SEQ ID NO:1019, SEQ ID NO:1020,
 SEQ ID NO:1021, SEQ ID NO:1022, SEQ ID NO:1023, SEQ ID
 2115 NO:1024, SEQ ID NO:1025, SEQ ID NO:1026, SEQ ID NO:1027,
 SEQ IDNO:1028, SEQ ID NO:1030, SEQ ID NO:1031, SEQ ID
 NO:1032, SEQ ID NO:1033, SEQ ID NO:1034, SEQ ID NO:1035,
 SEQ ID NO:1036, SEQ ID NO:1037, SEQ ID NO:1038, SEQ ID
 NO:1039, SEQ ID NO:1040, SEQ ID NO:1041,SEQ ID NO:1042,
 2120 SEQ ID NO:1043, SEQ ID NO:1044, SEQ ID NO:1045, SEQID
 NO:1046, SEQ ID NO:1047, SEQ ID NO:1048, SEQ ID NO:1049,
 SEQ ID NO:1050, SEQ ID NO:1051, SEQ ID NO:1052, SEQ ID
 NO:1053, SEQ ID NO:1054, SEQ ID NO:1056, SEQ ID NO:1057,
 SEQ ID NO:1058, SEQ ID NO:1059,SEQ ID NO:1061, SEQ ID
 2125 NO:1062, SEQ ID NO:1063, SEQ ID NO:1064, SEQID NO:1065,
 SEQ ID NO:1066, SEQ ID NO:1067, SEQ ID NO:1068, SEQ
 IDNO:1069, SEQ ID NO:1070, SEQ ID NO:1071, SEQ ID NO:
 1072, SEQ ID NO:1073, SEQ ID NO:1074, SEQ ID NO:1075.
 SEQ ID NO:1076, SEQ ID NO:1077, SEQ ID NO:1078, SEQ ID
 2130 NO:1079, SEQ ID NO:1080, SEQ ID NO:1081, SEQ ID NO:1082,

SEQ ID NO:1083, SEQ ID NO:1084, SEQ ID NO:1085, SEQ
 IDNO:1086, SEQ ID NO:1087, SEQ ID NO:1088, SEQ ID NO:
 1089, SEQ ID NO:1090, SEQ ID NO:1091, SEQ ID NO:1092,
 2135 SEQ ID NO:1094, SEQ ID NO:1095, SEQ ID NO:1096, SEQ ID
 NO:1097, SEQ ID NO:1098, SEQ ID NO:1099,SEQ ID NO:1100,
 SEQ ID NO:1101, SEQ ID NO:1102, SEQ ID NO:1103, SEQID
 NO:1104, SEQ ID NO:1105, SEQ ID NO:1106, SEQ ID NO:1107,
 SEQ ID NO:1108, SEQ ID NO:1109, SEQ ID NO:1110, SEQ ID
 NO:1111, SEQ ID NO:1112, SEQ ID NO:1113, SEQ ID NO:1114,
 2140 SEQ ID NO:1115, SEQ ID NO:1116,SEQ ID NO:1117, SEQ ID
 NO:1118, SEQ ID NO:1119, SEQ ID NO:1120, SEQID NO:1121,
 SEQ ID NO:1122, SEQ ID NO:1123, SEQ ID NO:1124, SEQ
 IDNO:1125, SEQ ID NO:1126, SEQ ID NO:1127, SEQ ID NO:
 1129, SEQ ID NO:1130, SEQ ID NO:1131, SEQ ID NO:1132,
 2145 SEQ ID NO:1133, SEQ ID NO:1134, SEQ ID NO:1135, SEQ ID
 NO:1136, SEQ ID NO:1137, SEQ ID NO:1138, SEQ ID NO:1139,
 SEQ ID NO:1140, SEQ ID NO:1141, SEQ ID NO:1142, SEQ
 IDNO:1143, SEQ ID NO:1144, SEQ ID NO:1145, SEQ ID NO:
 1146, SEQ ID NO:1147, SEQ ID NO:1148, SEQ ID NO:1149,
 2150 SEQ ID NO:1150, SEQ ID NO:1151, SEQ ID NO:1152, SEQ ID
 NO:1153, SEQ ID NO:1154, SEQ ID NO:1155,SEQ ID NO:1156,
 SEQ ID NO:1158, SEQ ID NO:1159, SEQ ID NO:1160, SEQID
 NO:1161, SEQ ID NO:1162, SEQ ID NO:1163, SEQ ID NO:1164,
 SEQ ID NO:1165, SEQ ID NO:1166, SEQ ID NO:1167, SEQ ID
 2155 NO:1168, SEQ ID NO:1169, SEQ ID NO:1170, SEQ ID NO:1171,
 SEQ ID NO:1172, SEQ ID NO:1173,SEQ ID NO:1174, SEQ ID
 NO:1175, SEQ ID NO:1176, SEQ ID NO:1177, SEQID NO:1178,
 SEQ ID NO:1179, SEQ ID NO:1180, SEQ ID NO:1181, SEQ
 IDNO:1182, SEQ ID NO:1183, SEQ ID NO:1184, SEQ ID NO:
 2160 1185, SEQ ID NO:1186, SEQ ID NO:1187, SEQ ID NO:1188,
 SEQ ID NO:1189, SEQ ID NO:1190, SEQ ID NO:1192, SEQ ID
 NO:1193, SEQ ID NO:1194, SEQ ID NO:1195, SEQ ID NO:1196,
 SEQ ID NO:1197, SEQ ID NO:1198, SEQ ID NO:1199, SEQ
 IDNO:1200, SEQ ID NO:1201, SEQ ID NO:1202, SEQ ID NO:

2165 1203, SEQ ID NO:1204, SEQ ID NO:1205, SEQ ID NO:1206,
 SEQ ID NO:1207, SEQ ID NO:1208, SEQ ID NO:1209, SEQ ID
 NO:1210, SEQ ID NO:1211, SEQ ID NO:1213,SEQ ID NO:1214,
 SEQ ID NO:1215, SEQ ID NO:1216, SEQ ID NO:1217, SEQID
 NO:1218, SEQ ID NO:1219, SEQ ID NO:1220, SEQ ID NO:1221,
 2170 SEQ ID NO:1222, SEQ ID NO:1223, SEQ ID NO:1224, SEQ ID
 NO:1225, SEQ ID NO:1226, SEQ ID NO:1227, SEQ ID NO:1228,
 SEQ ID NO:1229, SEQ ID NO:1230,SEQ ID NO:1231, SEQ ID
 NO:1232, SEQ ID NO:1233, SEQ ID NO:1234, SEQID NO:1235,
 SEQ ID NO:1236, SEQ ID NO:1237, SEQ ID NO:1238, SEQ
 2175 IDNO:1239, SEQ ID NO:1241, SEQ ID NO:1242, SEQ ID NO:
 1243, SEQ ID NO:1244, SEQ ID NO:1245, SEQ ID NO:1246,
 SEQ ID NO:1247, SEQ ID NO:1248, SEQ ID NO:1249, SEQ ID
 NO:1250, SEQ ID NO:1251, SEQ ID NO:1252, SEQ ID NO:1253,
 SEQ ID NO:1254, SEQ ID NO:1255, SEQ ID NO:1256, SEQ
 2180 IDNO:1257, SEQ ID NO:1259, SEQ ID NO:1260, SEQ ID NO:
 1261, SEQ ID NO:1262, SEQ ID NO:1263, SEQ ID NO:1264,
 SEQ ID NO:1265, SEQ ID NO:1266, SEQ ID NO:1267, SEQ ID
 NO:1268, SEQ ID NO:1269, SEQ ID NO:1270,SEQ ID NO:1271,
 SEQ ID NO:1272, SEQ ID NO:1273, SEQ ID NO:1275, SEQID
 2185 NO:1276, SEQ ID NO:1277, SEQ ID NO:1278, SEQ ID NO:1279,
 SEQ ID NO:1280, SEQ ID NO:1281, SEQ ID NO:1282, SEQ ID
 NO:1283, SEQ ID NO:1284, SEQ ID NO:1285, SEQ ID NO:1286,
 SEQ ID NO:1287, SEQ ID NO:1289,SEQ ID NO:1290, SEQ ID
 NO:1291, SEQ ID NO:1292, SEQ ID NO:1293, SEQID NO:1294,
 2190 SEQ ID NO:1295, SEQ ID NO:1296, SEQ ID NO:1297, SEQ
 IDNO:1298, SEQ ID NO:1299, SEQ ID NO:1300, SEQ ID NO:
 1301, SEQ ID NO:1303, SEQ ID NO:1304, SEQ ID NO:1305,
 SEQ ID NO:1306, SEQ ID NO:1307, SEQ ID NO:1308, SEQ ID
 NO:1310, SEQ ID NO:1311, SEQ ID NO:1312, SEQ ID NO:1313,
 2195 SEQ ID NO:1314, SEQ ID NO:1315, SEQ ID NO:1316, SEQ
 IDNO:1317, SEQ ID NO:1318, SEQ ID NO:1319, SEQ ID NO:
 1320, SEQ ID NO:1322, SEQ ID NO:1323, SEQ ID NO:1324,
 SEQ ID NO:1325, SEQ ID NO:1326, SEQ ID NO:1327, SEQ ID

NO:1328, SEQ ID NO:1330, SEQ ID NO:1331,SEQ ID NO:1332,
 2200 SEQ ID NO:1333, SEQ ID NO:1334, SEQ ID NO:1335, SEQ ID
 NO:1336, SEQ ID NO:1337, SEQ ID NO:1339, SEQ ID NO:1340,
 SEQ ID NO:1341, SEQ ID NO:1342, SEQ ID NO:1343, SEQ ID
 NO:1344, SEQ ID NO:1345, SEQ ID NO:1346, SEQ ID NO:1347,
 SEQ ID NO:1349, SEQ ID NO:1350,SEQ ID NO:1351, SEQ ID
 2205 NO:1352, SEQ ID NO:1353, SEQ ID NO:1354, SEQ ID NO:1355,
 SEQ ID NO:1356, SEQ ID NO:1357, SEQ ID NO:1358, SEQ
 IDNO:1360, SEQ ID NO:1361, SEQ ID NO:1362, SEQ ID NO:
 1363, SEQ ID NO:1364, SEQ ID NO:1365, SEQ ID NO:1367,
 SEQ ID NO:1368, SEQ ID NO:1369, SEQ ID NO:1370, SEQ ID
 2210 NO:1371, SEQ ID NO:1375, SEQ ID NO:1376, SEQ ID NO:1377,
 SEQ ID NO:1378, SEQ ID NO:1379, SEQ ID NO:1381, SEQ
 IDNO:1382, SEQ ID NO:1383, SEQ ID NO:1384, SEQ ID NO:
 1385, SEQ ID NO:1387, SEQ ID NO:1388, SEQ ID NO:1389,
 SEQ ID NO:1390, SEQ ID NO:1391, SEQ ID NO:1392, SEQ ID
 2215 NO:1393, SEQ ID NO:1395, SEQ ID NO:1396,SEQ ID NO:1397,
 SEQ ID NO:1398, SEQ ID NO:1399, SEQ ID NO:1400, SEQ ID
 NO:1402, SEQ ID NO:1403, SEQ ID NO:1404, SEQ ID NO:1405,
 SEQ ID NO:1406, SEQ ID NO:1407, SEQ ID NO:1409, SEQ ID
 NO:1410, SEQ ID NO:1412, SEQ ID NO:1413, SEQ ID NO:1414,
 2220 SEQ ID NO:1415, SEQ ID NO:1416,SEQ ID NO:1417, SEQ ID
 NO:1419, SEQ ID NO:1420, SEQ ID NO:1421, SEQ ID NO:1422,
 SEQ ID NO:1423, SEQ ID NO:1424, SEQ ID NO:1425, SEQ
 IDNO:1427, SEQ ID NO:1428, SEQ ID NO:1429, SEQ ID NO:
 1430, SEQ ID NO:1431, SEQ ID NO:1432, SEQ ID NO:1433,
 2225 SEQ ID NO:1434, SEQ ID NO:1435, SEQ ID NO:1437, SEQ ID
 NO:1438, SEQ ID NO:1439, SEQ ID NO:1440, SEQ ID NO:1441,
 SEQ ID NO:1442, SEQ ID NO:1444, SEQ ID NO:1445, SEQ
 IDNO:1446, SEQ ID NO:1447, SEQ ID NO:1448, SEQ ID NO:
 1449, SEQ ID NO:1451, SEQ ID NO:1452, SEQ ID NO:1453,
 2230 SEQ ID NO:1454, SEQ ID NO:1455, SEQ ID NO:1456, SEQ ID
 NO:1458, SEQ ID NO:1459, SEQ ID NO:1461,SEQ ID NO:1462,
 SEQ ID NO:1463, SEQ ID NO:1464, SEQ ID NO:1465, SEQ ID

NO:1466, SEQ ID NO:1468, SEQ ID NO:1469, SEQ ID NO:1470,
 SEQ ID NO:1472, SEQ ID NO:1474, SEQ ID NO:1475, SEQ ID
 2235 NO:1476, SEQ ID NO:1477, SEQ ID NO:1479, SEQ ID NO:1480,
 SEQ ID NO:1481, SEQ ID NO:1482,SEQ ID NO:1483, SEQ ID
 NO:1484, SEQ ID NO:1485, SEQ ID NO:1486, SEQID NO:1488,
 SEQ ID NO:1490, SEQ ID NO:1491, SEQ ID NO:1492, SEQ
 IDNO:1493, SEQ ID NO:1495, SEQ ID NO:1496, SEQ ID NO:
 2240 1497, SEQ ID NO:1498, SEQ ID NO:1500, SEQ ID NO:1502,
 SEQ ID NO:1503, SEQ ID NO:1504, SEQ ID NO:1505, SEQ ID
 NO:1507, SEQ ID NO:1509, SEQ ID NO:1512, SEQ ID NO:1513,
 SEQ ID NO:1514, SEQ ID NO:1515, SEQ ID NO:1517, SEQ
 IDNO:1518, SEQ ID NO:1519, SEQ ID NO:1521, SEQ ID NO:
 2245 1522, SEQ ID NO:1523, SEQ ID NO:1524, SEQ ID NO:1525,
 SEQ ID NO:1527, SEQ ID NO:1528, SEQ ID NO:1529, SEQ ID
 NO:1530, SEQ ID NO:1531, SEQ ID NO:1533,SEQ ID NO:1534,
 SEQ ID NO:1535, SEQ ID NO:1536, SEQ ID NO:1538, SEQID
 NO:1539, SEQ ID NO:1541, SEQ ID NO:1542, SEQ ID NO:1543,
 2250 SEQ ID NO:1544, SEQ ID NO:1546, SEQ ID NO:1548, SEQ ID
 NO:1550, SEQ ID NO:1552, SEQ ID NO:1554, SEQ ID NO:1556,
 SEQ ID NO:1557, SEQ ID NO:1559,SEQ ID NO:1560, SEQ ID
 NO:1561, SEQ ID NO:1562, SEQ ID NO:1564, SEQID NO:1565,
 SEQ ID NO:1567, SEQ ID NO:1568, SEQ ID NO:1570, SEQ
 2255 IDNO:1572, SEQ ID NO:1573, SEQ ID NO:1574, SEQ ID NO:
 1575, SEQ ID NO:1577, SEQ ID NO:1578, SEQ ID NO:1579,
 SEQ ID NO:1581, SEQ ID NO:1582, SEQ ID NO:1583, SEQ ID
 NO:1585, SEQ ID NO:1586, SEQ ID NO:1588, SEQ ID NO:1589,
 SEQ ID NO:1590, SEQ ID NO:1592, SEQ ID NO:1593, SEQ
 2260 IDNO:1595, SEQ ID NO:1597, SEQ ID NO:1598, SEQ ID NO:
 1600, SEQ ID NO:1602, SEQ ID NO:1606, SEQ ID NO:1608,
 SEQ ID NO:1609, SEQ ID NO:1610, SEQ ID NO:1611, SEQ ID
 NO:1613, SEQ ID NO:1614, SEQ ID NO:1616,SEQ ID NO:1618,
 SEQ ID NO:1620, SEQ ID NO:1621, SEQ ID NO:1623, SEQID
 2265 NO:1625, SEQ ID NO:1628, SEQ ID NO:1630, SEQ ID NO:1631,
 SEQ ID NO:1633, SEQ ID NO:1634, SEQ ID NO:1638, SEQ ID

NO:1641, SEQ ID NO:1642, SEQ ID NO:1644, SEQ ID NO:1645,
SEQ ID NO:1647, SEQ ID NO:1648,SEQ ID NO:1650, SEQ ID
NO:1651, SEQ ID NO:1653, SEQ ID NO:1654, SEQ ID NO:1656,
2270 SEQ ID NO:1657, SEQ ID NO:1659, SEQ ID NO:1661, SEQ
IDNO:1663, SEQ ID NO:1665, SEQ ID NO:1667, SEQ ID NO:
1671, SEQ ID NO:1674, SEQ ID NO:1676, SEQ ID NO:1678,
SEQ ID NO:1679, SEQ ID NO:1681, SEQ ID NO:1684, SEQ ID
NO:1686, SEQ ID NO:1687, SEQ ID NO:1689, SEQ ID NO:1692,
2275 SEQ ID NO:1693, SEQ ID NO:1695, SEQ ID NO:1697, SEQ
IDNO:1698, SEQ ID NO:1702, and SEQ ID NO:1703

, or (b) an amino acid sequence in the amino acid
sequences set forth in (a) in which several amino acids are
deleted, replaced or added.

2280 [0014]

The nucleic-acid molecule specific to enterohemorrhagic
pathogenic E. coli O-157:H7 of the present invention, a gene
included in the nucleic-acid molecule and a protein or a
polypeptide encoded by the gene are found by determining all
2285 nucleotide sequences on the chromosome of O-157:H7 SAKAI
and identifying a region and a nucleotide sequence specific to
O-157:H7 which are absent from nonpathogenic E. coli K-12.
The chromosomal nucleotide sequences of O-157:H7 determined
by the present invention have been registered on June 26, 2000,
2290 as Accession No. BA000007 in GenBank DDBJ.
[0015]

Furthermore, after the registration of the whole
chromosomal nucleotide sequence of O-157:H7 based on the
present invention, close similar nucleotide sequences to those
2295 of the present invention was registered on October 22, 2000
(GenBank/AE00517H). However, when these sequences were
registered, the sequences had two gaps and 2600 or more
characters other than AGCT (undetermined base). Thus the
sequences were imperfect. In addition, although the data
2300 thereof has been updated on September 25, 2001 and October 26,

2001, merely one gap sequence was determined and 2600 or more undetermined bases were remained.

[0016]

2305 In addition, as to obtained genetic information, homology search and prediction of predictive ORF and function thereof may be performed by comparison of the amino acid sequence to all sequence found in GenBank, DDBJ, SWISS-PROT and PIR database using an algorithm known in the art, for example, BLAST algorithm and the like.

2310 [0017]

The O-157:H7 specific polypeptides of the present invention are proteins or polypeptides having a character set forth in the tables described below. From the information of amino acid sequence, the polypeptides are classified to the following groups: 1) Proteins having unknown function etc., 2) Proteins which have unknown function, but have significant homology to that of other bacteria, 3) Proteins comprising Insertion Sequence; IS, 4) Proteins derived from phage, 5) Regulatory element, 6) Proteins relating to fimbriae, 7) Proteins relating to transportation of substance, 8) Proteins relating to synthesis of lipopolysaccharide, 9) Proteins relating to metabolism, 10) Proteins processing DNA/RNA, 11) Proteins relating to pathogenicity, 12) Other proteins.

[0018]

2325 List: polypeptides specific to enterohemorrhagic pathogenic *E. coli* O-157:H7

1) Proteins having a novel function

Sequence number: Hydrophobicity, The number of amino acids, Character such as function

2330 SEQ ID NO: 143: 0.610526, 39, novel

SEQ ID NO: 1438: -0.041667, 109, novel

SEQ ID NO: 1439: -0.505392, 817, an outer membrane usher protein precursor, similar to outer membrane usher protein precursors, for example, YehB [*Escherichia coli* K-12]

- 2335 gi|465572|sp|P33341|YEH#ECOLI (58% identity in the amino acids)
SEQ ID NO: 1440: -0.23304, 228, a putative fimbrial chaperone, similar to fimbrial chaperone, for example, YehC [Escherichia coli] gi|465573|sp|P33342|YEH#ECOLI (56% identity in 221 amino acids), GTG start
- 2340 SEQ ID NO: 1441: -0.121469, 178, a fimbrial major protein, similar to YehD [Escherichia coli] gi|465574|sp|P33343|YEH#ECOLI (26% identity in 177 amino acids), and similar to long polar fimbrial major proteins [Salmonella typhimurium] gi|1170815|sp|P43660|LPFA#SALTY (25% identity in 175 amino acids)
SEQ ID NO: 1442: -0.445877, 474, novel
SEQ ID NO: 1702: -0.448052, 78, similar to F plasmid CcdA protein (LetA protein) [Escherichia coli] gi|9507755|ref|NP#061421.1 (30% identity in 70 amino acids)
SEQ ID NO: 1703: 0.210577, 105, similar to F plasmid CcdB protein (LetB protein) [Escherichia coli] gi|9507756|ref|NP#061422.1 (35% identity in 104 amino acids)
- 2355 SEQ ID NO: 1663: -0.478836, 190, similar to YABP#ECOLI gi|2506583|sp|P39220 (38% identity in 168 amino acids)
SEQ ID NO: 1387: 0.060434, 370, a fimbrial protein, similar to putative putative fimbrial proteins, for example, [Escherichia coli] gi|538781|pir||B47152 (27% identity in the amino acids), and long polar fimbrial minor protein LpfE [Salmonella typhimurium] gi|1170819|sp|P43664|LPFE#SALTY (27% identity in 157 amino acids)
SEQ ID NO: 1388: -0.140816, 197, a putative fimbrial protein, similar to putative fimbrial protein YadK [Escherichia coli] gi|549488|sp|P37016|YADK#ECOLI (40% identity in 190 amino acids)
SEQ ID NO: 1389: -0.034826, 202, a putative fimbrial protein, similar to putative fimbrial protein YadL [Escherichia coli]

gi|549489|sp|P37017|YADL#ECOLI (41% identity in 192 amino acids)

2370 SEQ ID NO: 1390: -0.011828, 187, a putative fimbrial protein, similar to putative fimbrial-like protein YadM [Escherichia coli] gi|549490|sp|P37018|YADM#ECOLI (49% identity in 173 amino acids)

2375 SEQ ID NO: 1391: -0.387529, 867, similar to HTRE#ECOLI gi|1786332 (60% identity in 849 amino acids) [a putative outer membrane porin protein]

SEQ ID NO: 1392: -0.250623, 242, similar to ECPD#ECOLI gi|1786333 (60% identity in 239 amino acids) [a putative pilin

2380 chaperone]

SEQ ID NO: 1393: 0.058586, 199, similar to YADN#ECOLI gi|1786334 (39% identity in 195 amino acids) [a putative fimbrial-like protein]

SEQ ID NO: 979: -0.333674, 99, novel

2385 SEQ ID NO: 980: -0.245638, 150, novel

SEQ ID NO: 981: -0.622325, 216, novel, TTG start

SEQ ID NO: 982: -0.842466, 74, novel

SEQ ID NO: 983: -0.172956, 160, novel, similar to hypothetical 44.2kD protein YhhZ [Escherichia coli (strain K-12)]

2390 gi|1176284|sp|P46855|YHHZ#ECOLI (38% identity in 148 amino acids); and hemolysin-coregulated protein Hcp [Vibrio cholerae] gi|7467495|pir||T10891 (32% identity in 149 amino acids)

SEQ ID NO: 984: -0.448614, 470, novel

2395 SEQ ID NO: 985: -0.402126, 1036, novel, similar to IcmF protein [Legionella pneumophila] gi|7465644|pir||T18341 (20% identity in 1037 amino acids)

SEQ ID NO: 986: 0.637097, 63, novel, GTG start

SEQ ID NO: 987: -0.321591, 265, novel, GTG start

2400 SEQ ID NO: 988: -0.206311, 207, novel

SEQ ID NO: 989: 0.001619, 248, novel

SEQ ID NO: 990: -0.129036, 924, a putative ATP-dependent Clp

protease ATP-binding chain, similar to ATP-dependent Clp
 protease ATP-binding chain, for example, ClpB,
 2405 gi|7428220|pir||T07807, (40% identity in 753 amino acids)
 SEQ ID NO: 991: -0.11502, 254, novel [a putative membrane
 protein; IMP]
 SEQ ID NO: 992: -0.345146, 444, novel, its C-terminal part is
 similar to hypothetical protein z29f [Vibrio cholerae]
 2410 gi|3341578|emb|Caa13133.1|(51% identity in 104 amino acids)
 SEQ ID NO: 993: -0.308046, 175, novel [a hypothetical
 lipoprotein]
 SEQ ID NO: 994: -0.442019, 427, novel
 SEQ ID NO: 995: -0.298333, 361, novel
 2415 SEQ ID NO: 996: -0.314935, 617, novel
 SEQ ID NO: 997: -0.648175, 138, novel, similar to base plate
 proteins and acidolysozymes [coliphage T4]
 gi|137980|sp|P09425|VG25#BPT4 (34% identity in 62 amino
 acids) (at low level)
 2420 SEQ ID NO: 998: -0.380777, 464, novel, similar to
 hypothetical 54.5 kDa protein [Edwardsiella ictaluri]
 gi|2708666|gb|aaB92576.1|(41% identity in 461 amino acids)
 SEQ ID NO: 999: 0.109459, 75, novel
 SEQ ID NO: 1000: -0.366868, 167, novel, similar to a
 2425 hypothetical protein [Escherichia coli]
 gi|2920642|gb|aaC32477.1| (99% identity in 166 amino acids);
 and a hypothetical 19.5 kDa protein [Edwardsiella ictaluri]
 gi|2708667|gb|aaB92577.1|(32% identity in 148 amino acids)
 SEQ ID NO: 1001: -0.39593, 173, novel
 2430 SEQ ID NO: 1002: -0.16, 46, novel
 SEQ ID NO: 1003: -0.416269, 714, novel, similar to VgrG
 proteins, for example, [Escherichia coli strain ec11]
 gi|2920640|gb|aaC32475.1| (98% identity in 713 amino acids)
 SEQ ID NO: 1004: -0.707907, 1405, an Rhs protein, similar to
 2435 RhsH protein, for example, [Escherichia coli strain EC45]
 gi|2920634|gb|aaC32471.1| (92% identity in 1264 amino acids)

Appendix B: Hideo *et al.* Full Translation

SEQ ID NO: 1005: -0.704433, 204, novel, similar to YbcQ [Escherichia coli|gi|3025010|sp|P77234| (23% identity in 172 amino acids); and YibG [Escherichia coli|gi|418454|sp|P32106|YIBG#ECOLI (30% identity in 89 amino acids)

2440 SEQ ID NO: 1006: -0.305, 61, novel

SEQ ID NO: 1007: 1.333333, 97, novel [a hypothetical membrane protein: IMP]

2445 SEQ ID NO: 1008: -0.33836, 379, novel, similar to H repeat-associated proteins, for example, [Escherichia coli RhsB element|gi|140772|sp|P28912| (97% identity in 378 amino acids)

SEQ ID NO: 1009: -0.746417, 587, an Rhs protein, similar to Rhs core proteins, for example, RhsE [Escherichia coli|gi|2507113|sp|P24211|RHSE#ECOLI, TTG start

2450 SEQ ID NO: 1010: 0.701786, 57, novel, similar to N-terminal part of hypothetical protein, for example, ORF E2 in Rhs element [Escherichia coli|gi|2851489|sp|P31991|(92% identity in 56 amino acids)

2455 SEQ ID NO: 1011: -0.614943, 88, novel, similar to C-terminal part of hypothetical protein, for example, ORF E2 in Rhs element [Escherichia coli|gi|2851489|sp|P31991|(99% identity in 108 amino acids)

2460 SEQ ID NO: 1012: -0.31718, 391, novel, similar to H repeat-associated proteins, for example, [Escherichia coli RhsB element|gi|7465875|pir||E64898 (58% identity in 372 amino acids), GTG start

SEQ ID NO: 1094: -0.673765, 325, a putative integrase, similar to integrases, for example, [Shigella flexneri bacteriophage V|gi|2465477|gb|aaB72135.1| (88% identity in 305 amino acids)

2465 SEQ ID NO: 1095: -1.175308, 82, a transcription antitermination protein, partially similar to transcription antitermination protein N [Bacteriophage lambda|gi|73111|pir||VNBPL, (90% identity in 42 amino acids), may

2470

be disrupted

SEQ ID NO: 1096: -0.473644, 130, novel, similar to N-terminal part of hypothetical protein HP1334 [*Helicobacter pylori* (strain 26695)] gi|7464516|pir||F64686 (36% identity in 111 amino acids); and N-terminal part of hypothetical protein [*Neisseria meningitidis*] gi|6900422|emb|CAB72032.1|(31% identity in 113 amino acids)

SEQ ID NO: 1097: -0.28903, 238, a prophage repressor CI, similar to prophage repressor CI, for example, [Bacteriophage HK97] gi|6901592|gb|aaF31095.1|AF069529#8 (AF069529) (99% identity in 237 amino acids)

SEQ ID NO: 1098: -0.486364, 67, a Cro repressor, identical to regulatory protein Cro [phage lambda] gi|73101|pir||RCBP1; and similar to Cro protein, for example, [Bacteriophage HK97] gi|6901626|gb|aaF31129.1| (98% identity in 66 amino acid)

SEQ ID NO: 1099: -0.309278, 98, a regulatory protein cII, identical to regulatory protein cII [Bacteriophage lambda] gi|73106|pir||QCBP2L

SEQ ID NO: 1100: -0.622772, 203, a phage replication protein, similar to N-terminal part of phage replication protein, for example, O protein [Bacteriophage lambda] gi|75891|pir||ORBPL (88% identity in 163 amino acids), interrupted by frameshift

SEQ ID NO: 1101: -0.811764, 171, a phage replication protein, similar to C-terminal part of replication protein, for example, protein O [Bacteriophage lambda] gi|75891|pir||ORBPL (98% identity in 168 amino acids), interrupted by frameshift

SEQ ID NO: 1102: -0.002913, 104, a replication protein, its N-terminal part (amino acids at the position 1-21) is identical to replication protein P, for example, [Bacteriophage lambda] gi|75893|pir||PQBPL, probably disrupted

SEQ ID NO: 1103: -0.026894, 265, a putative tail fiber protein, partially similar to tail fiber proteins, for example, [Bacteriophage HK97] gi|6901608|gb|aaF31111.1| (AF069529)

- 2505 (42% identity in 155 amino acids); and similar to Sc/SvQ protein (DNA inversion product) [Escherichia coli plasmid p15B], for example, gi|96420|pir|S18690 (45% identity in 159 amino acids)
SEQ ID NO: 1104: -0.33198, 198, novel, similar to hypothetical
- 2510 proteins, for example, YcfA protein [Escherichia coli] gi|2506641|sp|P09153|YCFA#ECOLI (65% identity in 196 amino acids); Gp29 [Bacteriophage HK97] gi|6901609|gb|aaF31112.1|(66% identity in 192 amino acids); and T protein [Escherichia coli plasmid p15B]
- 2515 gi|96096|pir|S18684 (55% identity in 184 amino acids)
SEQ ID NO: 1105: -0.586394, 148, novel, similar to hypothetical proteins, for example, YfdK [Escherichia coli(strain K-12)] gi|3915468|sp|P77656|YFDK#ECOLI (68% identity in 144 amino acids)
- 2520 SEQ ID NO: 1106: -0.114706, 137, a putative tail fiber protein, similar to hypothetical proteins, for example, YfdL [Escherichia coli (strain K-12)] gi|2495635|sp|P76508|YFDL#ECOLI (52% identity in 67 amino acids); and putative tail fiber protein YcfE [Escherichia coli cryptic prophage e14]
- 2525 gi|7444558|pir|B64861 (51% identity in 45 amino acids)
SEQ ID NO: 1107: -0.234783, 185, a DNA-invertase, similar to DNA-invertases, for example, Pin [Escherichia coli] gi|72978|pir|JWEC (96% identity in 184 amino acids)
SEQ ID NO: 1108: -0.386771, 258, novel, similar to hypothetical
- 2530 protein [Deinococcus radiodurans (strain R1)] gi|7472205|pir|B75431 (32% identity in 249 amino acids)
SEQ ID NO: 1109: 0.763265, 50, novel
SEQ ID NO: 1110: 0.052227, 248, a putative transcription regulatory element, similar to transcription regulatory
- 2535 elements, for example, putative AraC-type regulatory protein YdeO gi|6176587|sp|P76135|YDEO#ECOLI (34% identity in 247 amino acids)
SEQ ID NO: 1111: -0.741026, 118, novel, similar to C-terminal

part of hypothetical protein, for example, [Escherichia coli
2540 insertion sequence IS2] gi|140808|sp|P19777|YI22#ECOLI
(77% identity in 113 amino acids), may be disrupted
SEQ ID NO: 1112: -0.510941, 394, a putative integrase, similar
to integrases, for example, [phage phi-R73]
gi|93827|pir|A42465 (61% identity in 388 amino acid)
2545 SEQ ID NO: 1113: -0.468841, 139, novel, GTG start
SEQ ID NO: 1114: -0.227805, 206, novel
SEQ ID NO: 1115: -0.045395, 153, novel
SEQ ID NO: 1116: -0.460952, 211, novel
SEQ ID NO: 1117: -0.462755, 197, novel, similar to
2550 hypothetical protein PFB0765w [malaria parasite]
gi|7494317|pir|E71606 (24% identity in 193 amino acids) (at
low level), TTG start
SEQ ID NO: 1118: -0.432979, 189, novel
SEQ ID NO: 1119: -0.854445, 91, a putative transcription
2555 activator, similar to Ogr family, for example, LsrS
[Rahnella aquatilis] gi|93826|pir|E42465 (41% identity and
65 amino acids); and delta protein [phage phi-R73]
gi|93826|pir|E42465 (36% identity in 76 amino acids)
SEQ ID NO: 1120: -0.291803, 184, a putative polarity
2560 suppression protein (amber mutation-suppression); similar to
Psu-like proteins, for example, Psu [Bacteriophage P4]
gi|1351414|sp|P05460|VPSU#BPP4 (30% identity in 166
amino acids)
SEQ ID NO: 1121: -0.4748, 251, a head size determination
2565 [protein], similar to head size determination proteins, for
example, Sid [phage phi-R73] gi|93821|pir|F42465 (22%
identity in 236 amino acids)
SEQ ID NO: 1122: -0.126744, 87, a putative DNA binding
protein, similar to hypothetical proteins, for example, putative
2570 DNABinding protein ORF88 [satellite phage P4]
gi|140147|sp|P12552|Y9K#BPP4 (65% identity in 82 amino
acids)

- SEQ ID NO: 1123: 0.40973, 186, a CI phage repressor, similar to CI repressors, for example, [Bacteriophage P4]
 2575 gi|1262833|emb|Caa35902.1| (67% identity in 115 amino acids)
- SEQ ID NO: 1124: -0.149315, 74, novel
- SEQ ID NO: 1125: 0.202804, 108, a putative copy number control protein, similar to orf106 [satellite phage P4]
 2580 gi|75896|pir||QQBPP4 (71% identity in 98 amino acids)
- SEQ ID NO: 1126: -0.193179, 778, a putative DNA primase, similar to DNA primases, for example, alpha gene product [satellite phage P4] gi|130905|sp|P10277|PRIM#BPP4 (72% identity in 770 amino acids)
- 2585 SEQ ID NO: 1127: -0.333019, 319, novel, similar to hypothetical protein 111401 [Synechocystis sp. (strain PCC 6803)] gi|7470073|pir||S74462 (21% identity in 206 amino acids), GTG start
- SEQ ID NO: 1451: 0.23625, 241, a putative oxidoreductase,
 2590 similar to oxidoreductases, for example, [Streptomyces coelicolor A3(2)] gi|6137024|emb|CAB59579.1| (55% identity in 237 amino acids)
- SEQ ID NO: 1452: 0.520652, 93, novel [hypothetical membrane protein; IMP]
- 2595 SEQ ID NO: 1453: 0.246154, 53, novel
- SEQ ID NO: 1454: -0.246667, 301, a putative transcription regulatory element (LysR family), similar to transcription regulatory elements, for example, [Xylella fastidiosa] gi|9106842|gb|aaF84577.1|AE003999#5 (40% identity in
 2600 290 amino acids)
- SEQ ID NO: 1455: -0.309788, 379, novel, similar to hypothetical protein, for example, [Pseudomonas aeruginosa] gi|732227|sp|Q01609|YODE#PSEAE (54% identity in 376 amino acids)
- 2605 SEQ ID NO: 1456: 0.996977, 398, a putative transporter protein, similar to transporters, for example, OpdE

- [*Pseudomonas* *aeruginosa*]
gi|400678|sp|Q01602|OPDE#PSEAE (60% identity in
396amino acid)
- 2610 SEQ ID NO: - : 0.215625, 97, novel
SEQ ID NO: 1577 : -0.388722, 134, novel, similar to
hypothetical proteins, for example, L0013 [*Escherichia coli*
O-157:H7 strain EDL933] gi|3414881|gb|aaC31492.1| (99%
identity in 133 amino acids), GTG start
- 2615 SEQ ID NO: 1578: 0.010435, 116, novel, similar to hypothetical
protein, for example, L0014 [*Escherichia coli* O-157:H7 strain
EDL933] gi|3288157|emb|Caa11510.1| (100% identity in 115
amino acids)
SEQ ID NO: 1579 : -0.445312, 513, novel, similar to
- 2620 hypothetical proteins, for example, L0015 [*Escherichia coli*
O-157:H7 strain EDL933] gi|3414883|gb|aaC31494.1| (100%
identity in 512 amino acids)
SEQ ID NO: - : -0.171316, 381, a putative NADH-dependent
flavin oxidoreductase, similar to YqiG [*Bacillus subtilis*]
- 2625 gi|1731054|sp|P54524|YQIG#BACSU (40% identity in 380
amino acids)
SEQ ID NO: 1495 : -0.089543, 307, novel, similar to
hypothetical proteins, for example, [*Escherichia coli* K-12]
gi|3183244|sp|P76049|YCJY#ECOLI (40% identity in 294
- 2630 amino acids) [in Tpx-Fnr intergenic region]
SEQ ID NO: 1496: -0.058117, 309, a putative transcription
regulatory element, similar to transcription regulatory
elements, for example, [*Escherichia coli*]
gi|2495398|sp|P75836|YCAN#ECOLI (38% identity in 291
- 2635 amino acids) [in DmsC-PflA intergenic region]
SEQ ID NO: 1497: -0.218644, 119, novel
SEQ ID NO: 1498: -0.25445, 192, a putative oxidoreductase,
similar to N-terminal part of oxidoreductase [aldo/keto
reductase family] (amino acids at the position 5-192/286), and
- 2640 similar to [*Thermotoga maritima*] gi|7431104|pir|A72308

- (59% identity in 185 amino acids)
 SEQ ID NO: - : -0.289344, 1418, a putative invasin, similar to putative membrane protein b1978 [Escherichia coli] gi|7466779|pir||D64962 (32% identity in 1352 amino acids)
 2645 and similar to vasins, for example, [Yersinia pestis] gi|726319|gb|aaA96352.1|(36% identity in 661 amino acids), and similar to intimins, for example, [Escherichia coli strain 4221] gi|1947048|gb|aa SEQ ID NO: acid B52913.1| [sic, gi|1947048|gb|aaB52913.1|] (30% identity in 874 amino acids)
 2650 SEQ ID NO: - : -0.170242, 290, a putative reductase, similar to reductases, for example, oxidoreductase, [Thermotoga maritima] gi|7431104|pir||A72308 (46% identity in 281 amino acids)
 SEQ ID NO: 1479: 0.107317, 83, novel, similar to hypothetical protein YaiU [Escherichia coli] gi|2495526|sp|P75700|YAIU#ECOLI (37% identity in 54 amino acids) [putative flagellin structural protein in HemB-sbmA intergenic region]
 2655 SEQ ID NO: 1480 -0.156319, 365, a putative adhesin, similar to high molecular weight adhesin, for example, HmWA [Haemophilus influenzae] gi|5929966|gb|aaD56660.1|AF180944#1 (19% identity in 199 amino acids)
 2660 SEQ ID NO: 1481: -0.088933, 254, novel
 2665 SEQ ID NO: 1482: -0.235772, 124, novel, similar to a part of hypothetical protein [Escherichia coli] gi|2506596|sp|P21514|YAHA#ECOLI (48% identity in 38 amino acids) ; and similar to regulatory elements, for example, BvgA [Bordetella bronchiseptica] gi|115157|sp|P16574|BVGA#BORPE (44% identity in 49 amino acids), GTG start
 2670 SEQ ID NO: 1483: 0.530909, 56, novel
 SEQ ID NO: 1484: -0.632692, 53, a putative fimbriaeregulatory protein, similar to invertase (partial), C-terminal part of type 1

- 2675 fimbriae regulatory proteins, for example, FimE [Escherichia coli K-12] gi|120167|sp|P04741|FIME#ECOLI (73% identity in 49 amino acids); and FimB [Escherichia coli] gi|729489|sp|P04742|FIMB#ECOLI (63% identity in 75 amino acids)
- 2680 SEQ ID NO: 1485 : -0.365069, 147, a putative fimbriae regulatory protein, invertase, similar to a part of type 1 fimbriae regulatory proteins, for example, FimB [Escherichia coli K-12] gi|729489|sp|P04742|FIMB#ECOLI (49% identity in 114 amino acids); and FimE [Escherichia coli] gi|120167|sp|P04741|FIME#ECOLI (42% identity in 113 amino acids), TTG start, probably interrupted
- 2685 SEQ ID NO: 1486: 1.684091, 45, novel
 SEQ ID NO: - : 0.114286, 50, novel
 SEQ ID NO: 1500: -0.450414, 1328, a putative adhesin, similar to AidA-I adhesin precursors, for example, [Escherichia coli plasmid F] gi|8918851|dbj|Baa97898.1| (45% identity in 1179 amino acids); similar to IgA1 protease homolog MisL [Salmonella typhimurium pathogenicity island SPI-3] gi|4324610|gb|aaD16954.1| (39% identity in 768 amino acids);
- 2695 and similar to VirG [Shigella flexneri] gi|96922|pir||A32247 (31% identity in 1014 amino acids)
- SEQ ID NO: 1502: -0.081707, 329, a putative sugar-binding protein, similar to sugar-binding proteins, for example, b1516 [Escherichia coli] gi|7466925|pir||G64905 (27% identity in 309 amino acids)
- 2700 SEQ ID NO: 1503: -0.030233, 87, a putative ABC transporter ATP-binding protein, similar to N-terminal part of ABC transporter ATP-binding protein, for example, [Streptomyces coelicolor A3(2)] gi|7479110|pir||T34924 (48% identity in 82 amino acid) [also to AraG-E.coli]
- 2705 SEQ ID NO: 1504: 0.144865, 371, a putative ABC transporter ATP-binding protein, similar to C-terminal part of sugar ABC transporter ATP-binding proteins, for example, [Bacillus

subtilis| gi|7404442|sp|P36947|RBSA#BACSU (36% identity
 2710 in 380 amino acids)
 SEQ ID NO: 1505: 0.929412, 324, a putative ABC transporter
 (permease) , similar to ABC transport system permeases, for
 example, RbsC [Bacillus subtilis| gi|7446897|pir||B69690
 (34% identity in 299 amino acids), and [Escherichia coli|
 2715 gi|400960|sp|P04984|RBSC#ECOLI (31% identity in 298
 amino acids)
 SEQ ID NO: - : 1.081132, 319, a putative ABCtransport
 system permease, similar to ABC transport system permeases,
 for example, RbsC [Escherichia coli| gi|78833|pir||C26304
 2720 (35% identity in 291 amino acids), and [Bacillus subtilis|
 gi|7446897|pir||B69690 (34% identity in 290 amino acids)
 SEQ ID NO: : -0.118928, 318, a putative transcription
 regulatory element, similar to araC-family transcription
 regulatory elements, for example, AdpA [Streptomyces
 2725 coelicolor A3(2)| gi|7544056|emb|CAB87229.1 (39% identity in
 311 amino acids)
 SEQ ID NO: 1606: -0.14084, 263, similar to YDDR#BACSU
 gi|7474951|pir||H69776 (47% identity in 259 amino acids)
 SEQ ID NO: 1360: -0.236079, 353, probably an ABC transporter
 2730 ATP-binding protein (probably ferric transport system), similar
 to ABC transporter ATP-binding proteins, for example, AfuC
 [Escherichia coli K-12| gi|2506109|sp|P37009|AFUC#ECOLI
 (94% identity in 352 amino acids)
 SEQ ID NO: 1361: 0.860259, 693, a putative ferrictransport
 2735 systempermease, similar to ferrictransport systempermeases,
 for example, AfuB [Actinobacillus pleuropneumoniae|
 gi|7387527|sp|Q44123|AFUB#ACTPL (66% identity in 671
 amino acids)
 SEQ ID NO: 1362 : -0.371429, 344, a putative
 2740 periplasmic-iron-binding protein, similar to
 periplasmic-iron-binding proteins, for example, AfuA
 [Actinobacillus pleuropneumoniae| gi|1469286|gb|aaB05032.1|

- (72% identity in 343 amino acids)
- SEQ ID NO: 1363 : 0.585714, 435, a putative regulatory
 2745 element, similar to hexosephosphate transport
 systemregulatory proteins, for example, UhpC [Escherichia coli
 K-12] gi|136770|sp|P09836|UHPC#ECOLI (53% identity in
 415 amino acids)
- SEQ ID NO: 1364: 0.329436, 514, a putative sensor histidine
 2750 protein kinase, similar to sensor protein kinases, for example,
 hexosephosphatetransport systemsensor protein UhpB
 [Escherichia coli K-12] gi|7429062|pir||RGECUB (35%
 identity in 497 amino acids)
- SEQ ID NO: 1365 : 0.151196, 210, a putative transcription
 2755 regulatory element (probably a response regulatory element),
 similar to transcription regulatory elements, for example,
 hexose phosphate transport system regulatory protein
 UhpA[Salmonella typhimurium]
 gi|136767|sp|P27667|UHPA#SALTY (49% identity in 202
 2760 amino acids); and UhpA [Escherichia coli]
 gi|136766|sp|P10940|UHPA#ECOLI (48% identity in 202
 amino acid)
- SEQ ID NO: - : 0.595302, 150, novel
- SEQ ID NO: 1625: -0.624948, 482, novel
- 2765 SEQ ID NO: 1697: -0.57125, 81, novel, similar to a part of
 hypothetical protein [Yersinia enterocolitica]
 gi|3511032|gb|aaC33681.1 (at the position 1-70 of 80 amino
 acids) (45% identity in 70 amino acids)
- SEQ ID NO: 1698: -0.341936, 94, novel, similar to hypothetical
 2770 protein (99 amino acids) [Yersinia pestis]
 gi|3822096|gb|aaC69816.1 (35% identity in 89 amino acids)
- SEQ ID NO: 1602: -0.638432, 524, novel
- SEQ ID NO: 1056: -0.363636, 452, a putative transporter (an
 outer membrane protein), similar to outer membrane
 2775 transporter proteins, for example, CyaE protein [Bordetella
 pertussis] gi|117799|sp|P11092|CYAE#BORPE (25% identity

- in 385 amino acids)
 SEQ ID NO: 1057 : 0.097741, 1462, novel, similar to
 hypothetical proteins, for example, [Synechocystis sp. strain
 2780 PCC 6803] gi|7469433|pir||S76109 (33% identity in 1384
 amino acids); similar to RTX protein [Aeromonas salmonicida]
 gi|6752871|gb|aaF27914.1|AF218037#1 (33% identity in 1384
 amino acids)
 SEQ ID NO: 1058 : -, 5292, novel, similar to
 2785 hypothetical proteins, for example, [Synechocystis sp. strain
 PCC 6803] gi|7469433|pir||S76109 (36% identity in 2014
 amino acids), and similar to RTX protein [Aeromonas
 salmonicida] gi|6752871|gb|aaF27914.1|AF218037#1 (36%
 identity in 2051 amino acids); hemagglutinin [Streptococcus
 2790 gordonii] gi|8885520|dbj|Baa97453.1| (35% identity in 2056
 amino acids), GTG start
 SEQ ID NO: 1059 : 0.082011, 707, a putative transporter,
 similar to transporter (ATP-binding proteins), for example,
 LktB [Actinobacillus
 2795 actinomycetemcomitans] gi|126357|sp|P23702|HLYB#ACTAC
 (26% identity in 690 amino acids)
 SEQ ID NO: - : -0.275448, 392, a putative transporter,
 similar to membrane fusion proteins, for example,
 [Sinorhizobium meliloti] gi|4689001|emb|CAB41456.1| (28%
 2800 identity in 372 amino acids)
 SEQ ID NO: 1559: -0.082857, 141, novel
 SEQ ID NO: 1560: 0.236364, 56, novel
 SEQ ID NO: 1561: -0.525147, 339, a putative adhesin/invasin,
 similar to surface protein [Xylella fastidiosa]
 2805 gi|9106565|gb|aaF84338.1|AE003982#11 (22% identity in 313
 amino acids); and putative adhesin/invasin [Neisseria
 meningitidis MC58] gi|7227256|gb|aaF42321.1| (23% identity
 in 337 amino acid)
 SEQ ID NO: 1562: -0.5825, 121, novel
 2810 SEQ ID NO: - : -0.746575, 74, novel, similar to a part of

- hypothetical protein YahH [Escherichia coli] gi|2495514|sp|P75690|YAHH#ECOLI (69% identity in 23 amino acids)
- SEQ ID NO: 1303: -0.35, 379, an H repeat-associated protein, similar to H repeat-associated protein in RhsB element [Escherichia coli] gi|140772|sp|P28912|YHH#ECOLI (97% identity in 378 amino acids)
- SEQ ID NO: 1304: -0.745946, 445, an Rhs protein, similar to putative Rhs proteintreptomyces coelicolor A3(2) gi|7321289|emb|CAB82067.1| (34% identity in 285 amino acids); and RhsE protein - E. coli gi|2507113|sp|P24211| (36% identity in 139amino acids), GTG start
- SEQ ID NO: 1305: -0.224444, 136, novel
- SEQ ID NO: 1306: -0.577477, 1617, an Rhs protein, similar to putative Rhs protein [Streptomyces coelicolor A3(2)] gi|7321289|emb|CAB82067.1| (30% identity in 857amino acids); and RhsH protein [Escherichia coli strain ec45] gi|2920634|gb|aaC32471.1| (25% identity in 919 amino acids)
- SEQ ID NO: 1307: -0.498693, 154, novel
- SEQ ID NO: 1308: -0.509795, 634, a putative Vgr protein, similar to Vgr protein, for example, [Escherichia coli strain ec11] gi|2920640|gb|aaC32475.1| (93% identity in 529 amino acid)
- SEQ ID NO: 1474: -0.281303, 354, similar to YBGO#ECOLI gi|1786935 (87% identity in 353 amino acids), but [having] different N-terminus
- SEQ ID NO: 1475: -0.419342, 244, similar to YBGP#ECOLI gi|1786936 (78% identity in 242 amino acids) [putative chaperone]
- SEQ ID NO: 1476: -0.430567, 724, similar to N-terminal part of YBGQ#ECOLI gi|1786937 (amino acids at the position 1-723/818) (84% identity in 723 amino acids) [putative outer membrane protein]
- SEQ ID NO: 1477: -0.026943, 194, similar to YBGD#ECOLI

2845 gi|1786938 (79% identity in 188 amino acids) [putative
fimbrial-like protein]
SEQ ID NO: 1275 : -0.0701, 302, a putative transcription
regulatory element, similar to transcription regulatory
elements, for example, glycine cleavage system transcription
2850 activator (gcv operon activator) - Escherichia coli
gi|417043|sp|P32064|GCVA#ECOLI (31% identity in 300
amino acids)
SEQ ID NO: 1276 : -0.4, 201, a putative cob(D)alamin
adenosyltransferase, similar to cob(D)alamin
2855 adenosyltransferases (corrinoid adenosyltransferases), for
example, [Escherichia coli]
gi|115148|sp|P13040|BTUR#ECOLI (67% identity in 200
amino acids)
SEQ ID NO: 1277 : -0.259636, 551, a putative fumarate
2860 hydratase, similar to fumarate hydratases, for example,
fumarate hydratase class I, aerobic (fumarase) - Escherichia
coli gi|120598|sp|P00923|FUMA#ECOLI (68% identity in 545
amino acids)
SEQ ID NO: 1278: 0.92183, 427, a putative transporter protein,
2865 similar to glutamate/aspartate transporter proteins (proton
glutamate symport proteins), for example, [Bacillus
stearothermophilus] gi|121467|sp|P24943|GLTT#BACST (38%
identity in 416 amino acids), and similar to
C4-dicarboxylate transporter proteins, for example, [Rhizobium
2870 l for example, uminosarum]
gi|231980|sp|Q01857|DCTA#RHILE (37% identity in 400
amino acids)
SEQ ID NO: 1279: -0.126667, 106, novel
SEQ ID NO: 1280 : -0.052632, 457, novel, similar to an
2875 unnamed protein product [Citrobacter amalonaticus]
gi|3184398|dbj|Baa28710.1| (93% identity in 284 amino acids)
SEQ ID NO: 1281 : -0.051816, 414, a 3-methylaspartate
ammonialyase (beta-methylaspartase), similar to

3-methylaspartate ammonia-lyases (beta-methylaspartases), for
 2880 example, [Citrobacter amalonaticus]
 gi|3184397|dbj|Baa28709.1| (93% identity in 413 amino
 acids); and [Clostridium tetanomorphum]
 gi|729971|sp|Q05514|MaaL#CLOTT (55% identity in 409
 amino acids)

2885 SEQ ID NO: 1282: -0.214345, 482, a probable glutamate
 mutase E (methylaspartate mutase E), similar to glutamate
 mutases, for example, [Citrobacter amalonaticus]
 gi|3184396|dbj|Baa28708.1| (90% identity in 481 amino acids),
 and [Clostridium tetanomorphum]
 2890 gi|729586|sp|Q05509|GLME#CLOTT (57% identity in 481
 amino acids)

SEQ ID NO: 1283: -0.058875, 463, a probable glutamate
 mutase L (methylaspartate mutase L), similar to glutamate
 mutase L (methylaspartate mutase L), for example,
 2895 [Clostridium tetanomorphum] gi|444421|prf||1907157C (32%
 identity in 449 amino acids)

SEQ ID NO: 1284: 0.061074, 150, a probable glutamate mutase
 S (methylaspartate mutase S), similar to glutamate mutase S
 (methylaspartate mutase S), for example, [Clostridium
 2900 Cochlearium] gi|7245512|pdb|1CCW|A (57% identity in 156
 amino acids)

SEQ ID NO: 1285: -0.278182, 56, novel
 SEQ ID NO: 1286: -0.114286, 141, novel
 SEQ ID NO: 1287: -0.327388, 315, novel

2905 SEQ ID NO: 928: -0.906945, 73, an excisionase, identical to
 excisionase [Bacteriophage HK022]
 gi|1722835|sp|P11683|VXIS#BP434; and similar to
 excisionase [Bacteriophage lambda]
 gi|139680|sp|P03699|VXIS#LAMBDA (98% identity in 72 amino
 2910 acids)

SEQ ID NO: 929: -0.565455, 56, novel, similar to hypothetical
 protein ORF55 [Bacteriophage 434] gi|801889|gb|aaA67903.1|

(98% identity in 55 amino acids)
 SEQ ID NO: 930: -0.0725, 41, novel, similar to hypothetical
 2915 protein ORF-91 [phage 434] gi|93720|pir||A27354 (82%
 identity in 28 amino acids)
 SEQ ID NO: 931: 0.247159, 177, novel [putative membrane
 protein: IMP]
 SEQ ID NO: 932: -0.605479, 74, novel, similar to C4-type zinc
 2920 finger proteins (TraR family), for example,
 gi|7649830|dbj|Baa94108.1| (98% identity in 73 amino acids)
 SEQ ID NO: 933: -0.346237, 94, novel, similar to hypothetical
 proteins, for example, [Bacteriophage 933W]
 gi|5881602|dbj|Baa84293.1| (97% identity in 93 amino acids):
 2925 and orf61 [Bacteriophage lambda] (95% identity in 46 amino
 acids)
 SEQ ID NO: 934: -0.079365, 64, novel, similar to hypothetical
 proteins, for example, [Bacteriophage VT2-Sa]
 gi|5881603|dbj|Baa84294.1| (96% identity in 61 amino acids),
 2930 and orf63 [Bacteriophage lambda] gi|508994|gb|aaA96567.1|
 (92% identity in 63 amino acids)
 SEQ ID NO: 935: -0.246667, 61, novel, similar to hypothetical
 protein, for example, [Bacteriophage 933W]
 gi|4585389|gb|aaD25417.1|AF125520#12 (95% identity in 60
 2935 amino acids) and orf60a [Bacteriophage lambda]
 gi|508995|gb|aaA96568.1| (93% identity in 60 amino acids)
 SEQ ID NO: 936: -0.359735, 227, an exonuclease, similar to
 exonucleases, for example, [Bacteriophage lambda]
 gi|119702|sp|P03697|EXO#LAMBDA (98% identity in 226 amino
 2940 acids)
 SEQ ID NO: 937: -1.293333, 61, novel, similar to NinE proteins,
 for example, [Bacteriophage 21] gi|4539480|emb|CAB39989.1|
 (95% identity in 60 amino acids)
 SEQ ID NO: 938: -0.675, 57, novel, similar to NinF proteins,
 2945 for example, [Bacteriophage 21] gi|4539481|emb|CAB39990.1|
 (92% identity in 56 amino acids), GTG start

SEQ ID NO: 939 : -1.100483, 208, novel, similar to NinG proteins, for example, [Bacteriophage 21] gi|4539482|emb|CAB39991.1| (95% identity in 204 amino acids)

2950 SEQ ID NO: 940 : -0.243891, 222, a serine/threonin proteinphosphatase, similar to serine/threonin proteinphosphatase, for example, [Bacteriophage lambda] gi|130792|sp|P03772|PP#LAMBD (95% identity in 221 amino acids)

2955 SEQ ID NO: 941 : -0.257367, 320, novel, [a putative outer membrane protein: OMP], similar to putative outer membrane protein [Helicobacter pylori (strain J99)] gi|7465285|pir|H71907 (19% identity in 297 amino acids)

2960 (at low level)

SEQ ID NO: 942: -0.396506, 230, antitermination, similar to antiterminators, for example, protein Q [Bacteriophage 82] gi|132277|sp|P13870|RegQ#BP82

SEQ ID NO: 943: 0.576577, 223, novel, [hypothetical membrane protein: IMP], similar to orf14 [Actinobacillus actinomycetemcomitans] gi|7592819|dbj|Baa94406.1| (29% identity in 228 amino acids); and TfpB protein [Moraxella bovis] gi|141258|sp|P20666|TFPB#MORBO (26% identity in 190 amino acids)

2965

SEQ ID NO: 944: -0.288636, 133, novel

2970 SEQ ID NO: 945: 0.109859, 72, an holin protein, holin proteins, for example, [Bacteriophage 933W] gi|4499808|emb|CAB39307.1| (92% identity in 71 amino acids)

SEQ ID NO: 946: -0.186061, 166, an endolysin (lysozyme), similar to endolysins (lysozyme), for example, R protein [Bacteriophage 21] gi|67436|pir|LZBP21 (93% identity in 165 amino acids)

2975

SEQ ID NO: 947: -0.409678, 156, novel, GTG start

SEQ ID NO: 948: -0.060294, 69, a ribosome protein L31-like

2980 protein, similar to hypothetical proteins, for example, ribosome

protein L31 homolog ykgM in intF-cacH intergenicregion
 [Escherichia coli K-12] gi|3025204|sp|P71302|YKGM#ECOLI
 (93% identity in 86amino acids), GTG start
 SEQ ID NO: 949: 0.736, 51, novel, GTG start

2985 SEQ ID NO: 950: 0.613043, 93, putative colicin immunity
 protein, similar to colicinimmunity proteins, for example,
 colicin E1 immunity protein
 gi|124395|sp|P02985|IMM1#ECOLI (25% identity in 107
 amino acid)

2990 SEQ ID NO: 951: -0.444172, 164, novel, [a putative membrane
 protein; IMP], similar to hypothetical protein MAL4P2.26
 [Plasmodium falciparum] gi|6562728|emb|CAB62867.1| (29%
 identity in 106 amino acids) (at low level)
 SEQ ID NO: 952: -0.572571, 701, novel

2995 SEQ ID NO: 953: -0.84, 71, novel
 SEQ ID NO: 954: -0.437433, 375, novel, similar to C-terminal
 part of hypothetical protein, for example, [Pseudomonas putida]
 gi|2995633|gb|aaC98738.1| (40% identity in 200 amino acids);
 and L0015 [Escherichia coli] gi|3414883|gb|aaC31494.1|

3000 (39% identity in 200 amino acids), GTG start
 SEQ ID NO: 955: -1.301176, 86, novel, similar to hypothetical
 protein, for example, orf29 [Escherichia coli]
 gi|6009405|dbj|Baa84864.1| (37% identity in 136 amino
 acids); and L0013 [Escherichia coli]

3005 gi|3414881|gb|aaC31492.1| (38% identity in 124 amino acids)
 SEQ ID NO: 956: -0.21966, 708, novel, similar to hypothetical
 proteins, for example, orf50 [Escherichia coli]
 gi|6009426|dbj|Baa84885.1| (71% identity in 106 amino
 acids); and L0014 [Escherichia coli]

3010 gi|3288157|emb|Caa11510.1| (64% identity in 116 amino
 acids)
 SEQ ID NO: 957: 0.07541, 123, novel, similar to hypothetical
 proteins, for example, L0015 [Escherichia coli]
 gi|3414883|gb|aaC31494.1| (61% identity in 503 amino acids)

- 3015 SEQ ID NO: 958: -0.213187, 92, novel, similar to hypothetical proteins, for example, 57.8 kD protein [Pseudomonas putida|gi|2496740|sp|P55630|Y4QI#RHISN (37% identity in 232 amino acids)
- 3020 SEQ ID NO: 959: -0.348958, 193, novel, similar to hypothetical protein, for example, 20.3K protein [Agrobacterium tumefaciens IS1131] gi|95090|pir||JC1151 (41% identity in 101 amino acids)
- SEQ ID NO: 960: -0.065414, 134, novel
- 3025 SEQ ID NO: 961: -0.125911, 248, immunity to R478 phage/colicin/tellurite resistance cluster, similar to TerW [plasmid R478] gi|1354147|gb|aaC44736.1| (99% identity in 155 amino acids)
- SEQ ID NO: 962: -0.134375, 129, novel
- 3030 SEQ ID NO: 963: -0.372477, 110, novel, similar to hypothetical proteins, for example, [Deinococcus radiodurans] gi|7472167|pir||B75302 (42% identity in 305 amino acids)
- SEQ ID NO: 964: -0.581686, 1022, novel, similar to hypothetical proteins, for example, [Streptomyces coelicolor A3(2)] gi|7472048|pir||A75302 (34% identity in 260 amino acids)
- 3035 SEQ ID NO: 965: -0.305505, 110, novel, similar to hypothetical proteins, for example, [Streptomyces coelicolor A3(2)] gi|8246803|emb|CAB92838.1| (45% identity in 97 amino acid)
- 3040 SEQ ID NO: 966: -0.476724, 233, novel, similar to hypothetical proteins, for example, [Serratia marcescens] gi|1695868|gb|aaB37122.1| (100% identity in 167 amino acids)
- SEQ ID NO: 967: -0.431156, 200, novel, hypothetical proteins, for example, [Serratia marcescens] gi|1695869|gb|aaB37123.1| (99% identity in 197 amino acids);
- 3045 and [Deinococcus radiodurans (strain R1)] gi|7471591|pir||F75301 (38% identity in 364 amino acids)
- SEQ ID NO: 968: 0.120465, 216, novel, similar to hypothetical proteins, for example, [Serratia marcescens]

gi|1695870|gb|aaB37124.1| (99% identity in 173 amino acid);
 3050 [Serratia marcescens] gi|1695871|gb|aaB37125.1| (98%
 identity in 53 amino acids); and [Deinococcus radiodurans]
 gi|7471522|pir|E75301 (28% identity in 286 amino acids)
 SEQ ID NO: 969: -0.357696, 1138, possible tellurium
 resistance, similar to TerZ protein, for example, [Serratia
 3055 marcescens] gi|6094454|sp|Q52353| (98% identity in 193
 amino acids)
 SEQ ID NO: 970: -0.31005, 200, a tellurium resistance, similar
 to TerA protein, for example, [Serratia marcescens]
 gi|5702379|gb|aaD47285.1|AF168355#3 (67% identity in 385
 3060 amino acids)
 SEQ ID NO: 971: -0.739041, 439, tellurite resistance, similar
 to TerB protein, for example, [Serratia marcescens]
 gi|950680|gb|aaA86848.1| (100% identity in 151 amino acids)
 SEQ ID NO: 972: -0.284314, 103, tellurium resistance, similar
 3065 to TerC protein, for example, [Serratia marcescens]
 gi|6226214|sp|Q52356|TERC#SERMA (100% identity in 346
 amino acids)
 SEQ ID NO: 973: -0.460736, 327, tellurium resistance, similar
 to terD protein, for example, [Serratia marcescens]
 3070 gi|6094448|sp|Q52357|TERD#SERMA (100% identity in 192
 amino acids)
 SEQ ID NO: 974: -0.541515, 331, possible tellurium resistance,
 identical to gi|7108482|gb|aaF36434.1|AF126104#3
 TLRB#ECOLI (100% identity in 191 amino acids); and similar to
 3075 TerE protein, for example, [Serratia marcescens]
 gi|6094449|sp|Q52358|TERE#SERMA (98% identity in 191
 amino acids)
 SEQ ID NO: 975: -0.394881, 294, novel
 SEQ ID NO: 976: 0.154545, 45, tellurium resistance, identical
 3080 to gi|7108481|gb|aaF36433.1|AF126104#2 TRLA#ECOLI
 (100% identity in 102 amino acids); and similar to TerF protein,
 for example, [Serratia marcescens]

Appendix B: Hideo *et al.* Full Translation

gi|7387491|gb|aaA86852.2| TERF#SERMA (94% identity in
102 amino acid)SEQ ID NO: 977: -0.360345, 233, novel, GTG
3085 start
SEQ ID NO: 1550: -0.338059, 671, an adhesin, similar to
Ihaadhesin [Escherichia coli O-157:H7 strain 86-24|
gi|7108480|gb|aaF36432.1|AF126104#1 IHA#ECOLI (99%
identity in 696 amino acids); and exogenous ferric siderophore
3090 receptor R4 [Escherichia coli strain CFT073|
gi|3661500|gb|aaC61730.1| gi|3661500|gb|aaC61730.1| (99%
identity in 669 amino acids)
SEQ ID NO: 1665: 0.638415, 165, novel, similar to a part of
hypothetical protein [Shigella flexneri|
3095 gi|5880472|gb|aaD54665.1|AF097520#3 (44 % identity in 40
amino acids)
SEQ ID NO: 1517: 0.82528, 448, novel, similar to C-terminal
part of ShiA [Shigella flexneri|
gi|5532447|gb|aaD44731.1|AF141323#2 (49% identity in 73
3100 amino acids); TTG start
SEQ ID NO: 1518: 0.075472, 107, novel
SEQ ID NO: 1519: -0.587221, 494, novel
SEQ ID NO: 1567: -0.283051, 414, novel, TTG start
SEQ ID NO: 1568: 0.021192, 152, novel, GTG start
3105 SEQ ID NO: - : 0.033871, 63, novel, TTG start
SEQ ID NO: 411: -0.575221, 340, novel
SEQ ID NO: 412: 0.496, 51, novel
SEQ ID NO: 413 : -0.713974, 824, a possible
glucosyl-transferase, similar to glucosyl-transferases, for
3110 example, [Salmonella typhi| gi|7467230|pir||T30292 (72%
identity in 366 amino acids)
SEQ ID NO: 414: 0.095238, 64, a putative ferric enterochelin
esterase (partial), similar to C-terminal part of ferric
enterochelin esterases, for example, [Salmonella enterica| gi|
3115 2738250|gb|aaC46181.1| (66% identity in 68amino acids), TTG
start

SEQ ID NO: 415: -0.280645, 63, a transposase, similar to transposases, for example, [Shigella boydii] gi|2197010|gb|aaB61273.1| (100% identity in 167 amino acids)

3120 SEQ ID NO: 416: -0.108911, 102, a possible repressor, similar to InsA protein, for example, [insertion sequence IS1F] gi|124915|sp|P19767|ISA2#ECOLI (98% identity in 91 amino acids), GTG start

SEQ ID NO: 417: -0.490164, 62, novel [putative membrane protein; IMP] SEQ ID NO: 418: -0.37, 51, novel

3125 SEQ ID NO: 419: -0.735659, 130, novel, GTG start

SEQ ID NO: 420: -0.62381, 43, novel, similar to sensor regulatory element protein HutT [Rhodobacter capsulatus] gi|1075537|pir|A49938 (33% identity in 97 amino acids) (at

3130 low level)

SEQ ID NO: 421: -0.882353, 52, novel

SEQ ID NO: 422: -0.729167, 73, novel

SEQ ID NO: 423: -0.036842, 96, transposase (OrfB), similar to transposases, for example, [insertion sequence IS629] gi|7443863|pir|T00315 (98% identity in 295 amino acids)

3135 SEQ ID NO: 424: -0.433333, 64, transposase (OrfA), similar to hypothetical proteins, for example, [Escherichia coli plasmid p O-157 insertion sequence IS629] gi|7444868|pir|T00241 (96% identity in 108 amino acids)

3140 SEQ ID NO: 425: -0.6728, 126, an HecB-like protein, its N-terminal-half part is similar to N-terminal part of hemolysin-activation protein HecB [Neisseria meningitidis MC58] gi|7227016|gb|aaF42103.1| (34% identity in 181 amino acids)

3145 SEQ ID NO: 426: -0.534445, 91, novel

SEQ ID NO: 427: -0.372341, 142, novel, similar to a part of tRNA-splicing endonuclease positive effector [fission yeast] gi|7493527|pir|T40065 (22% identity in 531 amino acids) (at

low level); and similar to hypothetical protein, for example,

3150 [Aquifexaeolicus] gi|7514764|pir|D70476 (24% identity in

271 amino acids) (at low level)
 SEQ ID NO: 428: -0.229139, 152, novel, TTG start
 SEQ ID NO: 429: -0.721212, 364, novel, similar to hypothetical
 proteins, for example, YbdN [Escherichia coli]
 3155 gi|3024984|sp|P77216|YBDN#ECOLI (58% identity in 396
 amino acids)
 SEQ ID NO: 430: -0.4, 249, novel, similar to hypothetical
 protein YbdM [Escherichia coli]
 gi|3024983|sp|P77174|YBDM#ECOLI (56% identity in 212
 3160 amino acids)
 SEQ ID NO: 431: -0.385547, 257, a transcription regulatory
 element, similar to PerC (BfpW) [Escherichia coli]
 gi|1172431|sp|P43475|PERC#ECOLI (25% identity in 83
 amino acids)
 3165 SEQ ID NO: 432: -0.49854, 138, novel, similar to
 exopolyphosphatase [Pseudomonas aeruginosa]
 gi|4200042|dbj|Baa74460.1| (32% identity in 56 amino acids)
 (at low level)
 SEQ ID NO: 433: -0.133074, 258, novel
 3170 SEQ ID NO: 434: 1.383019, 54, novel, its N-terminal part is
 similar to BfpM [Escherichia coli|gi|847983|gb|aaC44052.1|
 BFP#ECOLI (52% identity in 113 amino acids) ; its
 N-terminal part is similar to putative transposase [Vibrio
 cholerae|gi|7467523|pir|T09435 (55% identity in 68 amino
 3175 acids) ; and its C-terminal part is similar to a part of
 hypothetical protein [Escherichia coli O-157:H7]
 gi|7649865|dbj|Baa94143.1| (98% identity in 62 amino acids)
 SEQ ID NO: 435: 0.16, 46, novel, similar to hypothetical
 protein [Pseudomonas syringae|gi|1196744|gb|aaA88435.1|
 3180 (34% identity in 50 amino acids) (at low level)
 SEQ ID NO: 436: 0.065714, 71, novel, similar to hypothetical
 protein, for example, orf29 [Escherichia coli]
 gi|6009405|dbj|Baa84864.1| (40% identity in 131 amino
 acids); and L0013 [Escherichia coli]

Appendix B: Hideo *et al.* Full Translation

3185 gi|3414881|gb|aaC31492.1| (38% identity in 130 amino acids)
 SEQ ID NO: 437: -0.96087, 93, novel
 SEQ ID NO: 438: -0.462461, 326, novel, similar to hypothetical
 protein, for example, yfjP protein [Escherichia coli]
 gi|7449539|pir||B65042 (49% identity in 289 amino acids);
 3190 and yeeP protein [Escherichia coli]
 gi|2495624|sp|P76359|YEOP#ECOLI (95% identity in 183
 amino acids)
 SEQ ID NO: 439: -0.405691, 124, a putative adhesin, similar to
 outer membrane fluffing protein [Escherichia coli]
 3195 gi|7466262|pir||G64964 (68% identity in 927 amino acids);
 and similar to glyco protein [Escherichia coli strain H10407]
 gi|5305639|gb|aaD41751.1| (34% identity in 608 amino acids)
 (at low level); and similar to Adhesin AIDA-I precursor
 [Escherichia coli plasmid pIB6]
 3200 gi|543788|sp|Q03155|AIDA#ECOLI (23% identity in 678
 amino acids)
 SEQ ID NO: 440: -0.14065, 124, novel, similar to hypothetical
 protein YjDA [Escherichia coli]
 gi|731985|sp|P16694|YJDA#ECOLI (32% identity in 793
 3205 amino acids)
 SEQ ID NO: 441: 0.970589, 273, novel, similar to hypothetical
 protein YjcZ [Escherichia coli]
 gi|731984|sp|P39267|YJCZ#ECOLI (30% identity in 278 amino
 acids), GTG start
 3210 SEQ ID NO: 442: 0.125316, 80, novel
 SEQ ID NO: 443: 0.024615, 196, novel
 SEQ ID NO: 444: -0.242045, 617, novel, similar to hypothetical
 proteins, for example, YfjQ [Escherichia coli]
 gi|1723629|sp|P52132|YFJQ#ECOLI (73% identity in 271
 3215 amino acids); and YafZ [Escherichia coli]
 gi|2495487|sp|P77206|YAFZ#ECOLI (73% identity in 271
 amino acids)
 SEQ ID NO: 445: -0.965741, 109, novel, similar to hypothetical

proteins, for example, YafK [Escherichia coli]
 3220 gi|2495486|sp|P75676|YAFX#ECOLI (71% identity in
 144amino acids); and YfjX [Escherichia coli]
 gi|1723636|sp|P52139|YFJX#ECOLI (75% identity in 137
 amino acids)

SEQ ID NO: 446: -0.635945, 218, a putative DNA repair
 3225 protein (RadC family), similar to putative RadC family proteins,
 for example, YkfG [Escherichia coli]
 gi|3025218|sp|Q47685|YKFG#ECOLI (81% identity in 158
 amino acids); and YeeS [Escherichia
 coli|gi|3025155|sp|P76362|YEES#ECOLI (98% identity in 148
 3230 amino acids)

SEQ ID NO: 447: -0.957693, 105, novel, similar to hypothetical
 protein YeeT [Escherichia coli]
 gi|3025156|sp|P76363|YEET#ECOLI (97% identity in 73
 amino acids)

3235 SEQ ID NO: 448: 0.214754, 62, novel, similar to hypothetical
 proteins, for example, YeeU [Escherichia coli]
 gi|3025157|sp|P76364|YEEU#ECOLI (89% identity in
 118amino acids); and YfjZ [Escherichia coli]
 gi|1723638|sp|P52141|YFJZ#ECOLI (66% identity in 98 amino
 3240 acids), GTG start

SEQ ID NO: 449: -0.298065, 156, novel, similar to hypothetical
 proteins, for example, L0007 [Escherichia coli]
 gi|3414875|gb|aaC31486.1| (93% identity in 124 amino acids);
 YeeV [Escherichia coli| gi|3025158|sp|P76365|YEEV#ECOLI
 3245 (87% identity in 124 amino acids); and Ykfi [Escherichia coli]
 gi|3025213|sp|P77692|YKFI#ECOLI (58% identity in 112
 amino acids)

SEQ ID NO: 450: 0.945946, 38, novel, similar to hypothetical
 proteins, for example, L0008 [Escherichia coli]
 3250 gi|3414876|gb|aaC31487.1| (94% identity in 163 amino acids);
 and YeeW [Escherichia coli]
 gi|3025160|sp|P76366|YEEW#ECOLI (65% identity in 55

amino acids)

SEQ ID NO: 451: -0.110909, 56, novel, similar to hypothetical
 3255 proteins, for example, L0009 [Escherichia coli]
 gi|3414877|gb|aaC31488.1| (87% identity in 65 amino acids)

SEQ ID NO: 452: -0.405085, 178, novel, similar to hypothetical
 proteins, for example, L0010 [Escherichia coli]
 gi|3414878|gb|aaC31489.1| (81% identity in 111 amino acids);
 3260 ydiA [plasmid ColIb-P9] gi|4512489|dbj|Baa75138.1| (37%
 identity in 265 amino acids); and L0012 [Escherichia coli]
 gi|3414880|gb|aaC31491.1| (80% identity in 61 amino acids)

SEQ ID NO: 453: -0.335897, 79, novel

SEQ ID NO: 454: 0.984375, 65, a putative integrase, similar to
 3265 integrases, for example, [Escherichia coli prophage e14]
 gi|3024035|sp|P75969|INTE#ECOLI (46% identity in 372
 amino acids)

SEQ ID NO: 455: 0.088596, 115, a putative excisionase, similar
 to excisionase [bacteriophage P21]
 3270 gi|139674|sp|P27079|VXIS#BPP21 (31% identity in 73 amino
 acids)

SEQ ID NO: 456: 0.123529, 69, novel, GTG start

SEQ ID NO: 457: -0.905494, 92, novel, TTG start

SEQ ID NO: 458: -0.403175, 127, novel, similar to hypothetical
 3275 proteins, for example, YdfA [Escherichia coli]
 gi|140584|sp|P29008|YDFA#ECOLI (91% identity in 49 amino
 acids)

SEQ ID NO: 459: 0.010435, 116, a putative phage repressor,
 similar to repressor [Escherichia col Rac prophage]
 3280 gi|3025101|sp|P76062|RACR#ECOLI (91% identity in 158
 amino acids)

SEQ ID NO: 460: -0.445312, 513, novel, similar to YdaS
 [Escherichia coli] gi|3025102|sp|P76063|YDAS#ECOLI (84%
 identity in 94 amino acids)

3285 SEQ ID NO: 461: -0.04875, 81, novel, similar to YdaT
 [Escherichia coli] gi|3183265|sp|P76165|YDFX#ECOLI (31%

identity in 83 amino acids)
 SEQ ID NO: 462: -0.425233, 643, novel, similar to C-terminal
 part of replication termination protein DnaT (prepriming
 3290 protein I) [Escherichiacoli] gi|1361001|pir||S56589 (50%
 identity in 85 amino acids)
 SEQ ID NO: 463: -0.448868, 531, a putative replication protein,
 similar to replication proteins, for example , protein14
 [Bacteriophage phi-80] gi|137937|sp|P14814|VG14#BPPH8
 3295 (47% identity in 129 amino acids), GTG start
 SEQ ID NO: 464: 0.055688, 502, novel, similar to YdaW
 [Escherichia coli] gi|3025105|sp|P76066|YDAW#ECOLI (56%
 identity in 143 amino acids)
 SEQ ID NO: 465: -0.024348, 116, novel, GTG start
 3300 SEQ ID NO: 466: -0.331818, 89, novel, similar to Gp57
 [Bacteriophage N15] gi|7459176|pir||T13144 (69% identity in
 78 amino acids), GTG start
 SEQ ID NO: 467: -0.239801, 202, novel, similar to hypothetical
 protein, for example, [Bacteriophage VT2-Sa]
 3305 gi|5881670|dbj|Baa84361.1| (91% identity in 92 amino
 acids),GTG start
 SEQ ID NO: 468: -0.297006, 168, novel
 SEQ ID NO: 469: -0.163566, 130, novel, similar to hypothetical
 proteins, for example, Ea22 [Bacteriophage lambda]
 3310 gi|137663|sp|P03756|VE22#LAMBD (39% identity in 108
 amino acids), GTG start
 SEQ ID NO: 470: -0.442375, 860, novel
 SEQ ID NO: 471: -0.447707, 110, novel, its N-terminal part is
 similar to hypothetical proteins, for example, b2363
 3315 [Escherichia coli] gi|7451977|pir||H65009 (51% identity in 95
 amino acids), and its C-terminal part similar to hypothetical
 proteins, for example, [Bacteriophage 933W]
 gi|4585382|gb|aaD25410.1|AF125520#5 (43% identity in 75
 amino acids)
 3320 SEQ ID NO: 472: -0.339655, 233, novel

- SEQ ID NO: 473: -0.377251, 212, a prophage maintenance protein, similar to Hok/Gefffamily, for example, MokW [Bacteriophage 933W] gi|4585453|gb|aaD25481.1|AF125520#76 (90% identity in 70 amino acids)
- 3325 SEQ ID NO: 474: 0.057965, 227, novel, similar to QD1 [Bacteriophage N15] gi|2564084|gb|aaB81659.1| (31% identity in 64 amino acids)
- SEQ ID NO: 475: -0.939706, 69, novel, similar to b1560 [Escherichia coli] gi|1742555|dbj|Baa15259.1| (82% identity in 348 amino acids); and hypothetical protein A [phage P1] gi|732234|sp|Q06262|YORA#BPP1 (26% identity in 314 amino acids) (also to Orf19 (phi83)), GTG start
- 3330 SEQ ID NO: 476: -0.161714, 176, a putative crossover junction endodeoxyribonuclease, similar to Gp67 [Bacteriophage HK97] gi|6901639|gb|aaF31142.1| (59% identity in 110 amino acids); crossover junction endodeoxyribonucleases Rus [Escherichia coli cryptic lambdoid prophage DLP12] (41% identity in 107 amino acids); and gi|2507117|sp|P40116|RUS#ECOLI in (59% identity in 110 amino acids)
- 3340 SEQ ID NO: 477: -0.277615, 1158, a putative antitermination protein, similar to antitermination proteins, for example , proteinQ [Escherichia coli] gi|1742554|dbj|Baa15258.1| (39% identityin 273 amino acids)
- 3345 SEQ ID NO: 478: -0.279397, 200, novel, GTG start
- SEQ ID NO: 479: -0.658542, 440, novel, GTG start
- SEQ ID NO: 480: -0.259551, 90, novel, similar to hypothetical protein, for example, [Bacteriophage VT2-Sa] gi|5881634|dbj|Baa84325.1| (73% identity in 644 amino acids)
- 3350 SEQ ID NO: 481, ECs1125:1209796-1209978, -0.078333, 61, novel, similar to hypothetical protein [Bacteriophage 933W] gi|4499806|emb|CAB39305.1| (67% identity in 59 amino acids)
- SEQ ID NO: 482 -0.877248, 190, novel, similar to hypothetical proteins, for example, [Bacteriophage VT2-Sa]

- 3355 gi|5881635|dbj|Baa84326.1| (78% identity in 89 amino acids)
SEQ ID NO: 483: -0.436667, 61, a putative holin protein,
similar to holin proteins, for example, S protein [Bacteriophage
VT2-Sa] gi|5881636|dbj|Baa84327.1| (94% identity in 71
amino acids)
- 3360 SEQ ID NO: 245: -0.375688, 437, novel, similar to YdfR
[Escherichia coli] gi|3183262|sp|P76160|YDFR#ECOLI (47%
identity in 74 amino acids)
SEQ ID NO: 246: -0.447872, 95, a putative endolysin, similar
to endolysins, for example, R protein [Bacteriophage 933W]
- 3365 gi|4585422|gb|aaD25450.1|AF125520#45 (97% identity in 177
amino acid)
SEQ ID NO: 247: -0.294175, 104, a putative antirepressor
protein, identical to putative antirepressor protein
[Bacteriophage 933W]
- 3370 gi|4585423|gb|aaD25451.1|AF125520#46 ; and similar to
antirepressor protein Ant [BacteriophageP22]
gi|131843|sp|P03037|RANT#BPP22 (49% identity in 189
amino acids)
SEQ ID NO: 248: -0.781579, 115, an endopeptidase (host cell
3375 lysis), similar to endopeptidase, for example, Rz[Bacteriophage
VT2-Sa] gi|5881639|dbj|Baa84330.1|(80% identity in 155
amino acids)
SEQ ID NO: 249: -0.371015, 208, a lipoprotein Rz1precursor,
similar to lipoprotein Rz1 precursors, for example,
3380 [Bacteriophage 933W]gi|540738|pir||JN0750 (52% identity in
59 amino acids); [phage lambda]
gi|4585425|gb|aaD25453.1|AF125520#48 (76% identity in 59
amino acids)
SEQ ID NO: 250: -0.407368, 96, novel
- 3385 SEQ ID NO: 251: 0.416667, 73, novel, similar to hypothetical
protein [Bacteriophage VT2-Sa] gi|5881640|dbj|Baa84331.1|
(73% identity in 45 amino acids)
SEQ ID NO: 252: -0.590526, 96, novel

- SEQ ID NO: 253: -0.644516, 156, novel, similar to hypothetical
 3390 protein [Escherichia coli] gi|1778472|gb|aaB40755.1| (84%
 identity in 53 amino acids)
- SEQ ID NO: 254: -0.557587, 258, a putative DNase, similar to
 putative DNase [Bacteriophage phi-31]
 gi|1107475|emb|Caa62587.1| 28% identity in 85 amino acids)
- 3395 SEQ ID NO: 255: -0.615069, 74, a putative terminase small
 subunit, similar to terminasesmall subunit [Bacillus subtilis
 PBSX phage] gi|1722886|sp|P39785|XTMA#BACSU (42%
 identity in 57 amino acids), GTG start
- SEQ ID NO: 256: -0.595775, 72, a putative large terminase
 3400 subunit, similar to hypothetical proteins, for example, phage
 D3 terminase-like protein [Haemophilus influenzae]
 gi|6739656|gb|aaF27357.1|AF198256#11 (22% identity in 472
 amino acids); and similar to putative large terminase subunit
 [Bacteriophage A2] gi|3947452|emb|Caa07103.1| (25%
 3405 identity in 456 amino acids)
- SEQ ID NO: 257: -0.24127, 64, a putative major head
 protein/prohead protease, its N-terminal-half part is similar to
 putative prohead proteases, for example, Gp4
 [BacteriophageHK97] gi|1722780|sp|P49860|VP4#BPHK7
 3410 (28% identity in 136 amino acids); and its C-terminal-half part
 is similar to major head protein, for example, [Bacteriophage
 L5] gi|465114|sp|Q05223|VG17#BPML5 (23% identity in 280
 amino acids), GTG start
- SEQ ID NO: 258: -0.248333, 61, a putative portal protein,
 3415 similar to portal protein, for example, [Bacteriophage HK022]
 gi|6863114|gb|aaF30355.1|AF069308#3 (26% identity in 351
 amino acids)
- SEQ ID NO: 259: -0.338496, 227, novel, similar to a novel
 protein [Haemophilus influenzae]
 3420 gi|6739659|gb|aaF27360.1|AF198256#14 (71% identity in 21
 amino acids), GTG start
- SEQ ID NO: 260: -0.500383, 262, a putative head-tail adaptor,

similar to putative head-tail adaptors, for example,
 [Bacteriophage HK97] gi|6901597|gb|aaF31100.1| (45%
 3425 identity in 111 amino acids)
 SEQ ID NO: 261: -0.665942, 139, novel, similar to hypothetical
 phage protein, for example, Gp10 [Bacteriophage HK97]
 gi|6901598|gb|aaF31101.1| (75% identity in 148 amino acids)
 SEQ ID NO: 262: 0.008989, 90, novel, similar to Gp11
 3430 [Bacteriophage HK97] gi|6901599|gb|aaF31102.1| (49%
 identity in 113 amino acids)s
 SEQ ID NO: 263: -0.544444, 55, a putative major tail subunit,
 similar to major tail subunit [Bacteriophage HK97]
 gi|6901588|gb|aaF31091.1|AF069529#4 (66% identity in 234
 3435 amino acids)
 SEQ ID NO: 264: -0.273771, 123, a putative tail assembly
 chaperone, similar to tail assembly chaperon, for example, p14
 [Bacteriophage HK97] gi|6901600|gb|aaF31103.1| (62%
 identity in 124 amino acids)
 3440 SEQ ID NO: 265: -0.027711, 84, a putative tail protein [phage
 tail protein], similar to C-terminal part of Gp14 [Bacteriophage
 HK97] gi|6901601|gb|aaF31104.1| (60% identity in 90 amino
 acids), probably produced by translational frameshift
 SEQ ID NO: 266: -0.755556, 91, a putative tail length tape
 3445 measure protein (interrupted), similar to N-terminal part of
 tail length tape measure proteins, for example, [Bacteriophage
 HK97] gi|6901589|gb|aaF31092.1|AF069529#5 (81% identity
 in 137 amino acids)
 SEQ ID NO: 267: -0.881667, 61, a putative tail length tape
 3450 measure protein, similar to C-terminal part of tail length tape
 measure protein, for example, [Bacteriophage HK97]
 gi|6901589|gb|aaF31092.1|AF069529#5 (48% identity in 939
 amino acids), probably disrupted by frameshift
 SEQ ID NO: 268: 0.743396, 54, a putative minor tail protein.
 3455 similar to minor tail protein, for example, GpM [Bacteriophage
 lambda] gi|138845|sp|P03737|VMTM#LAMBDA (43% identity in

- 110 amino acids), GTG start
- SEQ ID NO: 269: -0.476879, 174, a putative minor tail protein, similar to minor tail protein, for example, GpL [Bacteriophage lambda] gi|138844|sp|P03738|VMTL#LAMBD (76% identity in 232 amino acids)
- 3460
- SEQ ID NO: 270: -0.315668, 218, a putative regulatory protein, similar to regulatory protein Mnt [Bacteriophage P22] gi|133138|sp|P03049|RMNT#BPP22 (34% identity in 73 amino acids)
- 3465
- SEQ ID NO: 271: -0.295775, 72, a putative antirepressor protein, its C-terminal part is similar to antirepressor proteins, for example, Ant[Bacteriophage P22] gi|131843|sp|P03037|RANT#BPP22 (84% identity in 71 amino acids), and its N-terminal part is similar to hypothetical phage proteins, for example, Gp30 [Bacteriophage N15] gi|7521545|pir|T13116 (35% identity in 175 amino acids)
- 3470
- SEQ ID NO: 272: -0.322449, 99, a putative tail assembly protein, similar to tail assembly proteins, for example, GpK [Bacteriophage lambda] gi|139638|sp|P03729|VTAK#LAMBD (86% identity in 196 amino acids)
- 3475
- SEQ ID NO: 273: -1.166667, 49, a putative tail assembly protein, similar to tail assembly protein, for example, GpI [Bacteriophage lambda] gi|139637|sp|P03730|VTAI#LAMBD (64% identity in 64 amino acids)
- 3480
- SEQ ID NO: 274: -0.734113, 300, a putative secreted effector protein, similar to secreted effector proteinopA [Salmonella dublin] gi|5669806|gb|aaD46479.1|AF121227#1 (31% identity in 587 amino acids)
- 3485
- SEQ ID NO: 275: -0.469565, 484, novel, its C-terminal part is similar to cytotoxic necrotizing factor type 2 [Escherichia coli] gi|1073353|pir|A55260 (31% identity in 244 amino acids) (its N-terminus is similar to a novel protein [P. falciparum] (at low level))
- 3490
- SEQ ID NO: 276: -0.447191, 90, novel

SEQ ID NO: 277: -0.883696, 93, novel [hypothetical membrane protein; IMP], similar to hypothetical protein, for example, b0362 [Escherichia coli] gi|7466098|pir|B64764(50% identity in 79 amino acids), [partially similar to hemin receptor precursor]

3495 SEQ ID NO: 278: -0.825352, 72, a transposase (OrfB) protein (insertionsequence IS2), similar to hypothetical protein, for example, [insertion sequence IS2] gi|140808|sp|P19777|YI22#ECOLI (98% identity in 301 amino acids), GTG start

3500 SEQ ID NO: 279: -, 79, novel, [putative transposase (OrfA)], similar to hypothetical protein [insertion sequence IS2] gi|140806|sp|P19776|YI21#ECOLI (100% identity in 53 amino acids)

3505 SEQ ID NO: 280 -0.735135, 149, novel, similar to hypothetical protein [Salmonella typhimurium LT2] gi|6960367|gb|aaF33527.1| (72% identity in 37 amino acids)

SEQ ID NO: 281: -0.217714, 176, novel

SEQ ID NO: 282: -1.381667, 61, novel, similar to Yop effector

3510 YopM [Yersinia enterocolitica] gi|4324334|gb|aaD16811.1| (25% identity in 171 amino acids), (also weakly to IpaH)

SEQ ID NO: 283: -0.215789, 58, novel, TTG start

SEQ ID NO: 284: -0.530738, 245, a putative integrase, similar to integrase, for example, [Shigella dysenteriae]

3515 gi|6759954|gb|aaF28112.1|AF153317#4 (31% identity in 389 amino acids)

SEQ ID NO: 285: -0.205833, 241, a putative DNA binding protein: similar to putative DNA binding protein (ORF88) [Bacteriophage P4] gi|140147|sp|P12552|Y9K#BPP4 (45% identity in 53 amino acids), GTG start

3520 SEQ ID NO: 286: -1.10199, 202, novel

SEQ ID NO: 287: -0.534375, 65, a putative cell division repressor, similar to cell division repressor Icd [enterobacteria phage P1] gi|4261623|gb|aaD13923.1|S61175#1 (42% identity

- 3525 in 45 amino acids)
 SEQ ID NO: 288: -0.325, 145, novel
 SEQ ID NO: 289: -0.088, 51, novel
 SEQ ID NO: 290: -0.079937, 320, novel
 SEQ ID NO: 291: -0.191011, 90, novel
- 3530 SEQ ID NO: 292: -0.281545, 635, novels
 SEQ ID NO: 293: -0.397973, 297, novel
 SEQ ID NO: 294: -0.965741, 109, novel
 SEQ ID NO: 295: 0.008475, 60, novel
 SEQ ID NO: 296: -0.431081, 149, novel
- 3535 SEQ ID NO: 297: 0.039437, 72, a putative single stranded
 DNA-binding protein, similar to single stranded DNA-binding
 proteins, for example, [Thermotoga maritima]
 gi|7439946|pir|H72354 (35% identity in 96 amino acids)
 SEQ ID NO: 298: -0.449153, 178, a putative transcription
- 3540 activator, similar to transcription activator of eaeA/bfpA, PerC
 (BfpW) [Escherichia coli] gi|1172431|sp|P43475|PERC#ECOLI
 (39% identity in 89 amino acids)
 SEQ ID NO: 299: -0.283069, 190, novel
 SEQ ID NO: 300: -0.520779, 155, a putative major head protein,
- 3545 similar to major head protein, for example, phage phi-C31
 gp36-like protein [Haemophilus influenzae]
 gi|6739663|gb|aaF27364.1|AF198256#18 (AF198256) (56%
 identity in 584 amino acids)
 SEQ ID NO: 301: 0.198361, 62, a putative prohead protease,
- 3550 similar to prohead proteases, for example, phage phi-C31
 gp35-like protein [Haemophilus influenzae]
 gi|6739662|gb|aaF27363.1|AF198256#17 (60% identity in 161
 amino acids)
 SEQ ID NO: 302: 0.183505, 98, a putative head portal protein,
- 3555 similar to head portal proteins, for example, phage phi-105
 ORF25-like protein [Haemophilus
 influenzae]gi|6739661|gb|aaF27362.1|AF198256#16 (63%
 identity in 403 amino acids)

SEQ ID NO: 303: -0.097403, 78, a putative head-tail adaptor,
 3560 similar to head-tail adaptors, for example, [Bacteriophage
 HK97] gi|6901597|gb|aaF31100.1| (47% identity in 112 amino
 acids)
 SEQ ID NO: 304: -0.730597, 269, novel, similar to hypothetical
 protein [Haemophilus influenzae]
 3565 gi|6739659|gb|aaF27360.1|AF198256#14 (45% identity in 98
 amino acids); and hypothetical protein 30 [Bacillus phage
 phi-105] gi|7459182|pir|T13519 (26% identity in 90 amino
 acids)
 SEQ ID NO: 305: -0.554049, 569, novel, similar to hypothetical
 3570 protein, for example, [Haemophilus influenzae]
 gi|6739658|gb|aaF27359.1|AF198256#13 (54% identity in 115
 amino acids)
 SEQ ID NO: 306: -0.527872, 715, novel
 SEQ ID NO: 307: -0.766567, 336, a putative terminase small
 3575 subunit, similar to hypothetical protein, genetic island 1
 [Haemophilus influenzae]
 gi|6739657|gb|aaF27358.1|AF198256#12 (64% identity in 112
 amino acids); and similar to putative terminase small subunit
 [Streptococcus thermophilus bacteriophage Sfi21]
 3580 gi|5230826|gb|aaD41028.1|AF112470#3 (29% identity in 98
 amino acids).
 SEQ ID NO: 308: -0.398762, 405, a putative terminase large
 subunit, similar to terminase large subunits, for example,
 [Haemophilus influenzae]
 3585 gi|6739656|gb|aaF27357.1|AF198256#11 (69% identity in 550
 amino acids), TTG start
 SEQ ID NO: 309: 0.25969, 130, novel
 SEQ ID NO: 310: -0.52549, 154, novel, GTG start
 SEQ ID NO: 311: -0.157219, 188, an integrase, similar to
 3590 integrases, for example, [Bacteriophage P21]
 gi|138558|sp|P27077|VINT#BPP21 (98% identity in 380 amino
 acids), (similar to lambda integrase)

Appendix B: Hideo *et al.* Full Translation

SEQ ID NO: 312: 0.063889, 217, an excisionase, similar to excisionases, for example, [Bacteriophage P21]
 3595 gi|139674|sp|P27079|VXIS#BPP21 (98% identity in 78 amino acids)

SEQ ID NO: 313: -0.793334, 646, a putative replication protein, similar to replication protein, for example, GpO [Bacteriophage lambda] gi|215150|gb|aaA96584.1| (69% identity in 261 amino acids)
 3600

SEQ ID NO: 314: -0.266292, 90, a replication protein, similar to replication proteins, for example, GpP [Bacteriophage lambda] gi|4499785|emb|CAB39284.1| (98% identity in 233 amino acids)

3605 SEQ ID NO: 315: -0.19875, 81, a putative Ren protein (protection from Rex-dependent exclusion), similar to Ren protein, for example, [Bacteriophage lambda] gi|139473|sp|P03761|VREN#LAMBDA (90% identity in 92 amino acids)

3610 SEQ ID NO: 316: 0.06375, 81, integral membrane drug resistance protein EmrE, similar to ethidium efflux protein EmrE (methyl viologen resistance protein C) [E. coli] gi|127565|sp|P23895|EMRE#ECOLI (98% identity in 110 amino acids), and belongs to the small multidrug resistance
 3615 (Smr) protein family

SEQ ID NO: 317: -0.018342, 568, novel, similar to hypothetical protein YbcK [Escherichia coli] gi|2495549|sp|P77698|YBCK#ECOLI (99% identity in 508 amino acids); and putative integrase [Bacteriophage A118]
 3620 gi|1196324|gb|aaB51416.1| (31% identity in 109 amino acids)

SEQ ID NO: 318: -0.248578, 423, novel, similar to hypothetical protein YbcN [Escherichia coli cryptic lambdoid prophage DLP12] gi|2495551|sp|Q47269|YBCN#ECOLI (92% identity in 151 amino acids), GTG start

3625 SEQ ID NO: 319: -0.218478, 93, novel, identical to NinE [Bacteriophage 82] gi|3024190|sp|Q37871|NINE#BP82

SEQ ID NO: 320: -0.159512, 206, novel, similar to YbcO [Escherichia coli cryptic prophage DLP12] gi|2495553|sp|Q47271|YBCO#ECOLI (97% identity in 96 amino acids); and Gp66 [Bacteriophage HK97] gi|6901638|gb|aaF31141.1| (68% identity in 95 amino acids)

3630 SEQ ID NO: 321: -0.289344, 245, a crossover junction endodeoxyribonuclease, similar to crossover junction endodeoxyribonucleases Rus, for example, [Escherichia coli bacteriophage 82] gi|2498868|sp|Q37873|RUS#BP82 (95% identity in 120 amino acids), GTG start

3635 SEQ ID NO: 322: -0.103759, 134, a putative antitermination protein, similar to antitermination protein, for example, Q[Bacteriophage 82] gi|132277|sp|P13870|RegQ#BP82 (98% identity in 229 amino acids)

3640 SEQ ID NO: 323: -0.622936, 219, a putative holin, similar to putative holin protein [Bacteriophage PS3] gi|3676074|emb|Caa09700.1| (72% identity in 103 amino acids), TTG start

3645 SEQ ID NO: 324: -0.662162, 149, a putative endolysin (lysozyme), similar to endolysins, for example, [Bacteriophage HK97] gi|6901642|gb|aaF31145.1| (95% identity in 158 amino acids)

[0019]

3650 2) Proteins which have novel function, but have significant homology

Sequence number: Hydrophobicity. The number of amino acids. Character such as function

3655 SEQ ID NO: 325: -0.109639, 84, a putative endopeptidase (host cell lysis), similar to hypothetical protein gp15 [Bacteriophage PS119] gi|3676087|emb|Caa09711.1| (83% identity in 155 amino acids): endopeptidases for example .[Bacteriophage lambda] gi|67522|pir|APBPML (59% identity in 153 amino acids)

3660 SEQ ID NO: 326: -0.749881, 422, a putative lipoprotein Rz1

precursor, lipoprotein Rz1 precursors, for example,
 [Bacteriophage lambda] (53% identity in amino acids)
 SEQ ID NO: 327: -0.631149, 2794, novel
 SEQ ID NO: 328: -0.122951, 62, novel [hypothetical
 3665 membrane protein; IMP]
 SEQ ID NO: 329: -0.232456, 115, novel
 SEQ ID NO: 330: 0.222857, 71, a putative terminase large
 subunit, similar to terminase large subunits, for example,
 [Bacteriophage WO] gi|6723224|dbj|Baa89621.1| (26%
 3670 identity in 641 amino acids); for example, [Bacteriophage N15]
 gi|7444579|pir||T13088 (25% identity in 630 amino acids)
 SEQ ID NO: 331: -0.754198, 132, novel
 SEQ ID NO: 332: -0.709589, 220, a putative portal protein,
 similar to putative portal protein [Wolbachia sp.
 3675 wKue|gi|6723246|dbj|Baa89642.1| (23% identity in 294 amino
 acids), GTG start
 SEQ ID NO: 333: -0.319445, 73, novel
 SEQ ID NO: 334: -0.243617, 95, a putative protease /scaffold
 protein, partially similar to ClpP proteases, for example,
 3680 [Bacteriophage D3] gi|5059251|gb|aaD38956.1| (35% identity
 in 218 amino acids); similar to putative scaffolding protein
 [Streptococcus thermophilus bacteriophage DT1]
 gi|4530143|gb|aaD21883.1| (30% identity in 201 amino acids)
 SEQ ID NO: 335: -0.664384, 74, novel, TTG start
 3685 SEQ ID NO: 336: -0.528708, 210, novel
 SEQ ID NO: 1570: -0.651901, 448, similar to minor tail proteins,
 for example, proteinZ [Bacteriophage N15]
 gi|7521219|pir||T13097 (52% identity in 192 amino acids);
 GpZ [Bacteriophage lambda]
 3690 gi|138849|sp|P03731|VMTZ#LAMBD (49% identity in 192
 amino acids)
 SEQ ID NO: 1030: 0.101176, 511, a putative minor tail
 component, similar to minor tail proteins, for example, protein
 U [Bacteriophage N15] gi|7444588|pir||T13098 (49% identity

3695 in 129 amino acids); GpU [Bacteriophage lambda]
gi|138847|sp|P03732|VMTU#LAMBDA (49% identity in 129
amino acids)
SEQ ID NO: 1031: -0.163804, 164, a major tail component,
similar to major tail proteins, for example, protein V
3700 [Bacteriophage N15] gi|7444589|pir|T13099 (62% identity in
244 amino acids); GpV [Bacteriophage lambda]
gi|138848|sp|P03733|VMTV#LAMBDA (55% identity in 246
amino acids)
SEQ ID NO: 1032: -0.270741, 271, a minor tail component,
3705 similar to minor tail proteins, for example, GpG [Bacteriophage
lambda] gi|138842|sp|P03734|VMTG#LAMBDA (33% identity in
109 amino acids)
SEQ ID NO: 1033: 0.038403, 264, a putative minor tail
component, similar to minor tail proteins, for example, GpT
3710 [Bacteriophage lambda] gi|138846|sp|P03735|VMTT#LAMBDA
(39% identity in 104 amino acids), probably produced by
translational frameshift
SEQ ID NO: 1034: -0.454546, 210, a putative tail length tape
measure protein precursor, similar to tail length tape measure
3715 protein precursors for example, GpH [Bacteriophage lambda]
gi|138843|sp|P03736|VMTH#LAMBDA (25% identity in 822
amino acids)
SEQ ID NO: 1035: -0.041442, 445, a putative minor tail
protein, similar to minor tail proteins for example, GpM
3720 [Bacteriophage lambda] gi|138845|sp|P03737|VMTM#LAMBDA
(55% identity in 108 amino acids)
SEQ ID NO: 1036: -0.442976, 841, a putative minor tail
protein, similar to minor tail proteins for example, GpL
[Bacteriophage lambda] gi|138844|sp|P03738|VMTL#LAMBDA
3725 (93% identity in 232 amino acids)
SEQ ID NO: 1037: -0.153648, 234, a putative tail assembly
protein, similar to tail assembly proteins for example, GpK
[Bacteriophage lambda] gi|139638|sp|P03729|VTAK#LAMBDA

- (97% identity in 199 amino acids)
- 3730 SEQ ID NO: 1038: 0.21129, 187, a putative tail assembly protein, similar to tail assembly proteins for example ,GpI [Bacteriophage lambda] gi|139637|sp|P03730|VTAI#LAMBD (80% identity in 215 amino acids)
- SEQ ID NO: 1039: -0.061353, 208, a putative host specificity
- 3735 protein, similar to host specificity proteins for example ,GpJ [Bacteriophage lambda] gi|138412|sp|P03749|VHSJ#LAMBD (88% identity in 1136 amino acids)
- SEQ ID NO: 1040: -0.166719, 1269, a putative outer membrane protein precursor, similar to outer membrane protein Lom
- 3740 precursors for example ,[prophage P-EibA] gi|7532789|gb|aaF63231.1|AF151091#2 (72% identity in 199 amino acids)
- SEQ ID NO: 1041: -0.41948, 540, a putative tail fiber protein, similar to tail fiber proteins for
- 3745 example ,[Bacteriophage 933W] gi|4585436|gb|aaD25464.1|AF125520#59 (67% identity in 277 amino acids)
- SEQ ID NO: 1042: 0.009016, 123, novel, similar to hypothetical proteins for example ,[Bacteriophage 933W]
- 3750 gi|4585437|gb|aaD25465.1|AF125520#60 (98% identity in 102 amino acids)
- SEQ ID NO: 1043: 0.422222, 190, novel, similar to hypothetical protein [Salmonella typhimurium LT2]
- gi|6960367|gb|aaF33527.1| (55% identity in 314 amino acids)
- 3755 SEQ ID NO: 1044: -0.17033, 183, novel
- SEQ ID NO: 1045: -0.29785, 94, novel
- SEQ ID NO: 1046: -0.139896, 387, novel
- SEQ ID NO: 1047: -0.09284, 853, novel
- SEQ ID NO: 1048: -0.12362, 327, novel, similar to secreted
- 3760 effector proteinopA,[Salmonella dublin] gi|5669806|gb|aaD46479.1|AF121227#1 (24% identity in 296 amino acids), similar to hypothetical proteins for

example ,YjBI [Escherichia coli]
 gi|418540|sp|P32690|YJBI#ECOLI (26% identity 183 amino
 3765 acids), weakly
 SEQ ID NO: 1049 : -0.341696, 284, novel [hypothetical
 membrane protein; IMP]
 SEQ ID NO: 1050: 0.074894, 236, a putative PTS transporter
 protein, similar to putative transporter proteins for
 3770 example ,SgaT [Escherichia coli]
 gi|2851673|sp|P39301|SGAT#ECOLI (38% identity in 440
 amino acids)
 SEQ ID NO:1051 : -0.083945, 219, a putative PTS system
 enzyme II, similar to phosphotransferase system enzymes IIBs
 3775 for example ,[Escherichia coli]
 gi|732028|sp|P39302|PTXB#ECOLI (28% identity in 99 amino
 acids)
 SEQ ID NO: 1052: 0.436468, 437, novel
 SEQ ID NO: 1053: -0.546947, 263, novel, GTG start
 3780 SEQ ID NO: 1054: -0.377489, 463, novel
 SEQ ID NO: 133: -0.3865, 401, unknown
 SEQ ID NO: 134 : -0.199834, 606, a putative integrase,
 similar to integrases for example ,[Bacteriophage HK022]
 gi|138560|sp|P16407|VINT#BPHK0 (27% identity in 321
 3785 amino acids)
 SEQ ID NO: 135: -0.420689, 146, novel
 SEQ ID NO: 136: -0.487755, 99, novel
 SEQ ID NO: 137 : -0.331236, 462, novel, similar to
 hypothetical proteins for example ,YdfD [Escherichia coli]
 3790 gi|140587|sp|P29010|YDFD#ECOLI (63% identity in 63 amino
 acids)
 SEQ ID NO: 138 : -0.780214, 188, a putative cell division
 inhibition, similar to dicB [Escherichia coli]
 gi|2507009|sp|P09557|DICB#ECOLI (54% identity in 62
 3795 amino acids)
 SEQ ID NO: 139: -0.17888, 787, novel, TTG start

Appendix B: Hideo *et al.* Full Translation

SEQ ID NO: 140: 0.226, 51, novel
 SEQ ID NO: 141: -0.445312, 513, novel
 SEQ ID NO: 142: 0.010435, 116, novel
 3800 SEQ ID NO: 143: -0.395489, 134, novel, similar to YdfB
 [Escherichia coli] gi|140585|sp|P29009|YDFB#ECOLI (100%
 identity in 41 amino acids)
 SEQ ID NO: 144: -0.538835, 104, novel, identical to YdfA
 [Escherichia coli] gi|140584|sp|P29008|YDFA#ECOLI (100%
 3805 identity in 51 amino acids)
 SEQ ID NO: 145: -0.684191, 273, novel, TTG start
 SEQ ID NO: 146: -0.275807, 249, novel, similar to
 hypothetical proteins for example ,yacB [plasmid ColIb-P9]
 gi|4512441|dbj|Baa75090.1| (35% identity in 92 amino acids)
 3810 SEQ ID NO: 147: -0.519277, 84, novel
 SEQ ID NO: 148: -0.448958, 97, a putative regulatory protein,
 similar to putative regulatory protein [Salmonella
 typhimurium] gi|7467281|pir|T03008 (30% identity in 108
 amino acids); DicA [Escherichia coli]
 3815 gi|118631|sp|P06966|DICA#ECOLI (27% identity in 108 amino
 acids)
 SEQ ID NO: 149: -0.025758, 67, novel
 SEQ ID NO: 150: 0.918487, 120, novel, similar to YdaT
 [Escherichia coli] gi|3025103|sp|P76064|YDAT#ECOLI (31%
 3820 identity in 141 amino acids)
 SEQ ID NO: 151: -0.246963, 429, novel
 SEQ ID NO: 152: 0.574468, 48, novel
 SEQ ID NO: 153: 0.214286, 92, a putative DNAREPLICATION
 protein, similar to DnaC homolog [Escherichia coli]
 3825 gi|7429001|pir|C64886 (79% identity in 248 amino acids);
 DnaC[Escherichia coli] gi|118715|sp|P07905|DNAC#ECOLI
 (48% identity in 242 amino acids)
 SEQ ID NO: 154: -0.016418, 68, novel, similar to
 gi|3025105|sp|P76066|YDAW#ECOLI (54% identity in 155
 3830 amino acids)

SEQ ID NO: 155: -0.025506, 248, novel

SEQ ID NO: 156: 0.022, 101, novel, similar to hypothetical proteins for example ,IroE[Salmonella enterica] gi|2738251|gb|aaC46182.1| (29% identity in 249 amino acids)

3835 SEQ ID NO: 157: -0.369811, 107, novel

SEQ ID NO: 158: -0.00581, 569, novel

SEQ ID NO: 159: -0.291558, 155, a putative prophage maintenance protein, similar to Hok/Gef family for example ,MokW [Bacteriophage 933W]

3840 gi|4585453|gb|aaD25481.1|AF125520#76 (92% identity in 65 amino acids)

SEQ ID NO: 160: -0.194196, 225, novel, similar to QD1 [Bacteriophage N15] gi|2564084|gb|aaB81659.1| (31% identity in 64 amino acids)

3845 SEQ ID NO: 161: -0.083415, 206, novel

SEQ ID NO: 162: -0.462832, 114, a putative crossover junction endodeoxyribonuclease, similar to Gp67 [Bacteriophage HK97] gi|6901639|gb|aaF31142.1| (60% identity in 113 amino acids); crossover junction endodeoxyribonuclease Rus

3850 [Escherichia coli cryptic prophage DLP12] gi|2507117|sp|P40116|RUS#ECOLI (40% identity in 115 amino acids)

SEQ ID NO: 163: 0.998039, 52, a putative antitermination protein, similar to bacteriophage antitermination proteins

3855 for example ,YbcQ [Escherichia coli cryptic prophage DLP12 gi|4585416|gb|aaD25444.1|AF125520#39 (77% identity in 124 amino acids)

SEQ ID NO: 164: -0.436782, 88, novel, similar to [hypothetical membrane protein] YpbD [Bacillus subtilis]

3860 gi|1730886|sp|P50730|YPBD#BACSU (30% identity in 128 amino acids)

SEQ ID NO: 165: -0.286022, 94, novel, similar to hypothetical protein [Bacteriophage P27] gi|8346569|emb|CAB93762.1| (97% identity in 49 amino acids)

- 3865 SEQ ID NO: 166: 0.757522, 114, a putative transcription regulatory element, similar to transcription regulatory elements for example ,YhiW [Escherichia coli] gi|586679|sp|P37638|YHIW#ECOLI (37% identity in 187 amino acids)
- 3870 SEQ ID NO: 167: 0.175785, 224, novel, similar to hypothetical proteins for example ,[Bacteriophage 933W] gi|4585419|gb|aaD25447.1|AF125520#42 (53% identity in 613 amino acids)
- SEQ ID NO: 168: -0.464706, 52, a transposase, identical to
- 3875 hypothetical protein [Escherichia coli plasmid p O-157 insertion sequence IS629] gi|7444868|pir||T00241 (100% identity in 116 amino acids)
- SEQ ID NO: 169: -0.152174, 254, a putative transposase, similar to transposases for example ,[Escherichia coli
- 3880 plasmid p O-157 insertion sequence IS629] gi|7443862|pir||T00240 (98% identity in 220 amino acids)
- SEQ ID NO: 170: -0.400502, 200, a putative transcription regulatory element, similar to PerC (BfpW) [Escherichia coli] gi|1172431|sp|P43475|PERC#ECOLI (47% identity in 87
- 3885 amino acids)
- SEQ ID NO: 171: -0.431915, 142, a lipoprotein Rz1 protein precursor, similar to Rz1 precursors for example ,[Bacteriophage 933W] gi|4585425|gb|aaD25453.1|AF125520#48(98% identity in 61
- 3890 amino acids); [Bacteriophage lambda] gi|540738|pir||JN0750(70% identity in 61 amino acids)
- SEQ ID NO: 172: -0.121552, 117, a endopeptidase (host cell lysis), similar to endopeptidases for example ,[Bacteriophage VT2-Sa] gi|5881639|dbj|Baa84330.1| (88% identity in 154
- 3895 amino acids)
- SEQ ID NO: 173: -0.561452, 538, novel
- SEQ ID NO: 174: -0.275207, 243, novel
- SEQ ID NO: 175: -0.345833, 121, a host cell lysis, similar to

endolysins for example ,[Bacteriophage H-19B]
 3900 gi|4335686|gb|aaD17382.1| (94% identity in 177 amino acids)
 SEQ ID NO: 176: -0.521101, 110, novel
 SEQ ID NO: 177: -0.46, 156, novel
 SEQ ID NO: 178: -0.444527, 403, novel
 SEQ ID NO: 179: -0.033648, 319, a holin protein (host cell
 3905 lysis), similar to holin proteins for example ,[Bacteriophage
 VT2-Sa] gi|5881636|dbj|Baa84327.1| (91% identity in 69
 amino acids)
 SEQ ID NO: 180: 0.066393, 245, novel, GTG start
 SEQ ID NO: 181 : -0.292064, 127, novel, similar to
 3910 hypothetical proteins for example ,L0013 [Escherichia coli
 O-157:H7 strain EDL933] gi|3414881|gb|aaC31492.1|(99%
 identity in 133 amino acids)
 SEQ ID NO: 182 : -0.271985, 258, novel, identical to
 hypothetical proteins for example ,L0014 [Escherichia coli
 3915 O-157:H7 strain EDL933] gi|3414882|gb|aaC31493.1|(100%
 identity in 115 amino acids)
 SEQ ID NO: 183 : -0.112369, 381, novel, similar to
 hypothetical proteins for example ,L0015 [Escherichia coli
 O-157:H7 strain EDL933] gi|3414883|gb|aaC31494.1|(100%
 3920 identity in 512 amino acids)
 SEQ ID NO: 184: -0.165341, 353, a putative terminase small
 subunit, similar to C-terminal part of terminase small subunits
 for example ,[Bacteriophage N15]
 gi|2507082|sp|P31061|NOHA#ECOLI(46% identity in 75
 3925 amino acids), GTG start, probably disrupted by IS insertion
 SEQ ID NO: 185: -0.206736, 194, a terminase large subunit,
 similar to terminase large subunits for
 example ,[Bacteriophage 21]
 gi|2851579|sp|P36693|TERL#BPP21 (91% identity in 637
 3930 amino acids)
 SEQ ID NO: 186: -0.392375, 342, a portal protein, similar to
 portal proteins for example ,GP4 [Bacteriophage P21]

- gi|549295|sp|P36272|VG04#BPP21 (98% identity in 530 amino acids)
- 3935 SEQ ID NO: 187: -0.188742, 152, a head-tail preconnector protein, similar to head-tail preconnector proteins for example ,Gp5 [Bacteriophage P21] gi|549296|sp|P36273|VG05#BPP21 (97% identity in 501 amino acids), GTG start
- 3940 SEQ ID NO: 188: 0.734105, 347, a head decoration protein, similar to head decoration proteins for example ,Gpshp [Bacteriophage P21] gi|549437|sp|P36275|VSH#BPP21 (95% identity in 115 amino acids)
- SEQ ID NO: 189: -0.317188, 193, a possible major head protein,
- 3945 similar to N-terminal part of major head proteins for example ,Gp7 [Bacteriophage P21] gi|547612|sp|P36270|HEAD#BPP21(95% identity in 88 amino acids)
- SEQ ID NO: 190: -0.249738, 192, novel
- 3950 SEQ ID NO: 191: 0.297015, 68, a putative tail component, similar to minor tail proteins for example ,GpG [Bacteriophage lambda] gi|138842|sp|P03734|VMTG#LAMBD (68% identity in 143 amino acids)
- SEQ ID NO: 192: -0.083333, 103, a putative minor tail
- 3955 component, similar to minor tail protein GpG-T [Bacteriophage lambda] gi|7429179|pir|TLBPTL (72% identity in 124 amino acids), probably produced by translational frameshift
- SEQ ID NO: 193: 0, 75, a tail length determinant, similar to tail length tape measure proteins for example ,GpH
- 3960 [Bacteriophage lambda] gi|138843|sp|P03736|VMT#LAMBD (77% identity in 859 amino acids)
- SEQ ID NO: 194: -0.427011, 697, a minor tail component, similar to minor tail proteins for example ,GpM [Bacteriophage lambda] gi|138845|sp|P03737|VMTM#LAMBD (82% identity in 109 amino acids)
- 3965 SEQ ID NO: 195: 0.565, 41, a minor tail component, similar to

- minor tail proteins for example ,GpL [Bacteriophage lambda]
gi|138844|sp|P03738|VMTL#LAMBD (76% identity in 232
amino acids)
- 3970 SEQ ID NO: 196: 0.101111, 91, a tail assembly protein,
similar to tail assembly proteins for example ,GpK
[Bacteriophage lambda] gi|139638|sp|P03729|VTAK#LAMBD
(84% identity in 196 amino acids)
SEQ ID NO: 197: -0.5, 51, a tail assembly protein, similar to
- 3975 tail assembly proteins for example ,GpI [Bacteriophage
lambda] gi|139637|sp|P03730|VTAI#LAMBD (68% identity in
224 amino acids)
SEQ ID NO: 198: -1.1875, 65, novel
SEQ ID NO: 199: -0.140541, 75, a copper/zinc superoxide
- 3980 dismutase, similar to copper/zinc-superoxide dismutases for
example ,[Salmonella typhimurium]
gi|2462699|emb|Caa73588.1| (58% identity in 175 amino
acids)
SEQ ID NO: 200: -0.113333, 91, a putative host specificity
- 3985 protein, similar to host specificity proteins for example ,GpJ
[Bacteriophage lambda] gi|138412|sp|P03749|VHSJ#LAMBD
(65% identity in 1156 amino acids)
SEQ ID NO: 201: -0.59375, 65, a putative outer membrane
- 3990 protein, similar to Lom outer membrane proteins for
example ,[prophage P-EibA]
gi|7532789|gb|aaF63231.1|AF151091#2 (68% identity in 199
amino acids)
SEQ ID NO: 202: 0.147917, 49, a putative tail fiber protein,
- 3995 similar to putative tail fiber proteins for
example ,[Bacteriophage 933W]
gi|4585436|gb|aaD25464.1|AF125520#59 (38% identity in 370
amino acids)
SEQ ID NO: 203: -0.707843, 103, novel, similar to
- 4000 hypothetical protein [Bacteriophage 933W]
gi|4585437|gb|aaD25465.1|AF125520#60 (93% identity in 129

amino acids); similar to C-terminal part of putative tail protein [933W] gi|4585436|gb|aaD25464.1|AF125520#59(93% identity in 89 amino acids)

SEQ ID NO: 204: 0.03369, 375, novel, GTG start

4005 SEQ ID NO: 205: -0.295604, 92, a putative secreted effector protein, similar to EspF proteins for example ,[Escherichia coli strain E2348/69] gi|2865308|gb|aaC38400.1| (37% identity in 87 amino acids); L0016 - Escherichia coli gi|3414884|gb|aaC31495.1| (38% identity in 126 amino acids)

4010 SEQ ID NO: 206: -0.495808, 168, novel, partially similar to avirulence protein A [Pseudomonas syringae] gi|114726|sp|P11437|AVRA#PSESG (46% identity in 56 amino acids)

SEQ ID NO: 207: -0.350549, 92, a putative integrase, identical to integrase [Bacteriophage 933W] gi|4585378|gb|aaD25406.1|AF125520#1, but [having] different start; similar to integrases for example ,[Escherichia coli rac prophage] gi|6166234|sp|P76056|INTR#ECOLI (42% identity in 408 amino acids)

4020 SEQ ID NO: 208: 0.199342, 153, a putative excisionase, identical to putative excisionase [Bacteriophage 933W] gi|4585379|gb|aaD25407.1|AF125520#2

SEQ ID NO: 209: 0.463492, 64, novel, identical to hypothetical protein [Bacteriophage 933W] gi|4585380|gb|aaD25408.1|AF125520#3

4025 SEQ ID NO: 210: -0.033136, 170, novel, identical to hypothetical protein [Bacteriophage 933W] gi|4585381|gb|aaD25409.1|AF125520#4, but [having] different start

4030 SEQ ID NO: 211: -0.402415, 208, novel, identical to hypothetical protein [Bacteriophage 933W] gi|4585382|gb|aaD25410.1|AF125520#5; similar to hypothetical protein [Bacteriophage 933W]

Appendix B: Hideo *et al.* Full Translation

4035 gi|4585455|gb|aaD25483.1|AF125520#78 (50% identity in 80 amino acids)

SEQ ID NO: 212 : -0.577922, 78, novel, identical to hypothetical protein [Bacteriophage 933W]

4040 gi|4585383|gb|aaD25411.1|AF125520#6 (100% identity in 95 amino acids)

SEQ ID NO: 213 : 0.356338, 72, novel, identical to hypothetical protein [Bacteriophage 933W]

gi|4585384|gb|aaD25412.1|AF125520#7 (100% identity in 72 amino acids), GTG start

4045 SEQ ID NO: 214 : -0.410847, 296, novel, identical to hypothetical protein [Bacteriophage 933W]

gi|4585385|gb|aaD25413.1|AF125520#8 (100% identity in 95 amino acids), GTG start

4050 SEQ ID NO: 215 : -0.942593, 109, novel, identical to hypothetical protein [Bacteriophage VT2-Sa]

gi|5881600|dbj|Baa84291.1| (100% identity in 155 amino acids)

SEQ ID NO: 216 : -0.260656, 245, novel, identical to hypothetical protein [Bacteriophage 933W]

4055 gi|4585386|gb|aaD25414.1|AF125520#9 (100% identity in 257 amino acids);, similar to hypothetical proteins for example ,[Bacteriophage 933W]

gi|4585455|gb|aaD25483.1|AF125520#78 (95% identity in 157 amino acids), GTG start

4060 SEQ ID NO: 217 : -0.421638, 172, novel, similar to C4-type zinc finger proteins (TraR family) for example ,gi|4585456|gb|aaD25484.1|AF125520#79 (79% identity in 73 amino acids)

4065 SEQ ID NO: 218 : -0.312093, 646, novel, identical to hypothetical protein [Bacteriophage 933W], but [having] different start; similar to orf61 [Bacteriophage lambda]

gi|508993|gb|aaA96566.1| (93% identity in 46 amino acids)

Appendix B: Hideo *et al.* Full Translation

SEQ ID NO: 219 : -0.186957, 47, novel, identical to
4070 hypothetical protein [Bacteriophage VT2-Sa]
gi|5881603|dbj|Baa84294.1| (100% identity in 63 amino
acids); similar to orf63 [Bacteriophage lambda]
gi|508994|gb|aaA96567.1| (90% identity in 61 amino acids)
SEQ ID NO: 220 : -0.418537, 411, novel, identical to
4075 hypothetical protein [Bacteriophage VT2-Sa]
gi|5881604|dbj|Baa84295.1|, but [having] different start;
similar to orf60a [Bacteriophage lambda]
gi|508995|gb|aaA96568.1| (96% identity in 60 amino acids)
SEQ ID NO: 221: -0.531132, 213, a exonuclease, similar to
4080 exonuclease [Bacteriophagelambda] gi|2981722|pdb|1AVQ|A
(98% identity in 226 amino acids)
SEQ ID NO: 222: -0.137079, 90, a recombination protein Bet,
identical to Bet [Bacteriophage VT2-Sa]
gi|5881606|dbj|Baa84297.1| (100% identity in 261 amino
4085 acids); similar to Bet [Bacteriophage lambda]
gi|137511|sp|P03698|VBET#LAMBDA (99% identity in 261
amino acids)
SEQ ID NO: 223: -0.533645, 215, a host-nuclease inhibitor
protein Gam, similar to Gam proteins for
4090 example ,[Bacteriophage lambda]
gi|138128|sp|P03702|VGAM#LAMBDA (97% identity in 138
amino acids)
SEQ ID NO: 224: -0.435294, 52, a Kil protein, identical to kil
[Bacteriophage VT2-Sa] gi|5881608|dbj|Baa84299.1|; similar
4095 to kill proteins for example ,[Bacteriophage lambda]
gi|138622|sp|P03758|VKIL#LAMBDA (98% identity in 89 amino
acids)
SEQ ID NO: 225: -0.714458, 167, a regulatory protein cIII
(antitermination), identical to cIII [Bacteriophage lambda]
4100 gi|133366|sp|P03044|RPC3#LAMBDA (100% identity in 54
amino acids)
SEQ ID NO: 226: 0.126027, 74, a single strandbinding protein

Appendix B: Hideo *et al.* Full Translation

Ea10, identical to Ea10 [Bacteriophage VT2-Sa]
 gi|5881610|dbj|Baa84301.1| (100% identity in 122 amino
 4105 acids); similar to Ea10 [Bacteriophage lambda]
 gi|137630|sp|P03757|VE10#LAMBD (99% identity in 122
 amino acids)
 SEQ ID NO: 227 : -0.575177, 142, novel, identical to
 hypothetical protein [Bacteriophage VT2-Sa]
 4110 gi|5881612|dbj|Baa84303.1| (100% identity in 83 amino acids)
 SEQ ID NO: 228: -1.413333, 61, a putative anti-termination
 N protein, identical to N protein [Bacteriophage VT2-Sa]
 gi|5881613|dbj|Baa84304.1|, but [having] different start;
 similar to N proteins for example ,[Bacteriophage 933W]
 4115 gi|4585397|gb|aaD25425.1|AF125520#20 (42% identity in 90
 amino acids)
 SEQ ID NO: 229: -0.125172, 291, novel
 SEQ ID NO: 230: -0.297787, 950, novel
 SEQ ID NO: 231 : -0.469647, 795, novel, identical to
 4120 hypothetical protein [Bacteriophage VT2-Sa]
 gi|5881614|dbj|Baa84305.1| (100% identity in 173 amino
 acids)
 SEQ ID NO: 232: -0.370764, 302, a putative cI repressor
 protein, similar to cI [Bacteriophage lambda]
 4125 gi|133353|sp|P03034|RPC1#LAMBD (70% identity in 208
 amino acids)
 SEQ ID NO: 233: 0.007584, 357, a putative regulatory protein,
 identical to hypothetical protein [Bacteriophage VT2-Sa]
 gi|5881616|dbj|Baa84307.1|; similar to c2 [Bacteriophage L]
 4130 gi|1469215|emb|Caa63999.1| (42% identity in 49 amino acids)
 SEQ ID NO: 234: 0.418519, 55, a regulatory protein CII,
 identical to CII protein [Bacteriophage VT2-Sa]
 gi|5881617|dbj|Baa84308.1| (100% identity in 98 amino
 acids); similar to CII proteins for example ,[Enterobacteria
 4135 phage HK022] gi|631957|pir|S42398 (96% identity in 98
 amino acids)

SEQ ID NO: 235 : -0.554044, 273, novel, identical to hypothetical protein [Enterobacteria phage HK022] gi|632160|pir||S42399 (100% identity in 48 amino acids):
4140 similar to orf48 [Bacteriophage P22] gi|871503|emb|Caa55155.1| (85% identity in 48 amino acids)
SEQ ID NO: 236: -0.290062, 162, a endopeptidase (host cell lysis), similar to endopeptidases for example ,[Bacteriophage lambda] gi|119368|sp|P00726|ENPP#LAMBD (97% identity in
4145 153 amino acids)
SEQ ID NO: 237: -0.084177, 159, a lipoprotein Rz1 precursor, similar to Rs1 precursors for example ,[Bacteriophage lambda] gi|540738|pir||JN0750 (96% identity in 60 amino acids)
SEQ ID NO: 238: -0.384931, 74, novel, similar to Bor protein precursors for example ,[Bacteriophage lambda] gi|137520|sp|P26814|VBOR#LAMBD (98% identity in 97 amino acids)
4150
SEQ ID NO: 239 : -0.322581, 125, novel, similar to hypothetical proteins for example ,YbcV [Escherichia coli] gi|2495556|sp|P77598|YBCV#ECOLI (98% identity in 150 amino acids)
4155
SEQ ID NO: 240: -0.276613, 125, novel, identical to YbcW [Escherichia coli] gi|2495557|sp|P75720|YBCW#ECOLI
SEQ ID NO: 241 : 0.049693, 164, novel, similar to
4160 hypothetical proteins for example ,[Escherichia coli] gi|1778472|gb|aaB40755.1| (98% identity in 64 amino acids)
SEQ ID NO: 242: -0.307692, 66, a terminase small subunit, similar to terminase smallsubunits for example ,Nul [Bacteriophage lambda] gi|139026|sp|P03707|TERS#LAMBD
4165 (97% identity in 181 amino acids)
SEQ ID NO: 243: -0.415, 281, a putative terminase large subunit, similar to terminase large subunits for example , protein A [Bacteriophage lambda] gi|137616|sp|P03708|TERL#LAMBD (99% identity in 641
4170 amino acids), GTG start

- SEQ ID NO: - : 0.61519, 80, a head-to-tail joining protein, similar to head-to-tail joining proteins for example ,GpW [Bacteriophage lambda] gi|138415|sp|P03727|VHTJ#LAMBD (98% identity in 68 amino acids)
- 4175 SEQ ID NO: 485: -0.691397, 373, a putative portal protein, similar to portal proteins for example ,GpB [Bacteriophage lambda] gi|138762|sp|P03710|VMCB#LAMBD (98% identity in 533 amino acids)
- 4180 SEQ ID NO: 486: -0.496629, 90, a minor capsid protein, similar to minor capsid proteins for example , protein C [Bacteriophage lambda] gi|137565|sp|P03711|VCAC#LAMBD (97% identity in 439 amino acids), GTG start, containing Nu3-homolog
- 4185 SEQ ID NO: 487: -0.65931, 146, a major capsid protein, similar to major capsid proteins for example ,GpD [Bacteriophage lambda] gi|137566|sp|P03712|VCAD#LAMBD(99% identity in 110 amino acids)
- 4190 SEQ ID NO: 488: 0.03027, 186, a putative major capsid protein, similar to major capsid proteins for example ,GpE [Bacteriophage lambda] gi|116752|sp|P03713|HEAD#LAMBD (98% identity in 341 amino acids)
- 4195 SEQ ID NO: 489: -0.356579, 77, a DNA packaging protein, similar to DNA packaging proteins for example ,GpFI [Bacteriophage lambda] gi|139324|sp|P03709|VPF1#LAMBD (98% identity in 132 amino acids)
- 4200 SEQ ID NO: 490: -0.53038, 159, a minor capsid protein, similar to minor capsid proteins for example ,GpFII [Bacteriophage lambda] gi|137575|sp|P03714|VCF2#LAMBD(94% identity in 117 amino acids), GTG start
- SEQ ID NO: 491: -0.797196, 108, a minor tail protein, similar to minor tail proteins for example ,GpZ [Bacteriophage lambda] gi|138849|sp|P03731|VMTZ#LAMBD (98% identity in 192 amino acids)

- 4205 SEQ ID NO: 492: -0.397163, 142, a minor tail protein, similar to minor tail proteins for example ,GpU [Bacteriophage lambda] gi|138847|sp|P03732|VTU#LAMB (100% identity in 131 amino acids)
- SEQ ID NO: 493: -0.69942, 346, a major tail protein V,
- 4210 similar to major tail proteins for example ,GpV [Bacteriophage lambda] gi|138848|sp|P03733|VMTV#LAMB (95% identity in 246 amino acids)
- SEQ ID NO: 494: -0.687309, 198, a minor tail protein, similar to minor tail proteins for example ,GpG [Bacteriophage lambda] gi|138842|sp|P03734|VMTG#LAMB (96% identity in 140 amino acids)
- 4215 SEQ ID NO: 495: -0.404622, 239, a putative minor tail protein, similar to minor tail proteins for example ,GpT [Bacteriophage lambda] gi|138846|sp|P03735|VMTT#LAMB (99% identity in 144 amino acids), probably produced by translational frameshift
- SEQ ID NO: 496: -0.494286, 106, a tail length tape measure protein precursor, similar to tail length tape measure protein precursors for example ,GpH [Bacteriophage lambda] gi|138843|sp|P03736|VMTH#LAMB (96% identity in 849 amino acids)
- 4225 SEQ ID NO: 497: -0.175, 101, a minor tail protein, similar to minor tail proteins for example ,GpM [Bacteriophage lambda] gi|138845|sp|P03737|VMTM#LAMB (94% identity in 109 amino acids)
- 4230 SEQ ID NO: 498: -0.355238, 106, a minor tail protein, similar to minor tail proteins for example ,GpL [Bacteriophage lambda] gi|138844|sp|P03738|VMTL#LAMB (98% identity in 232 amino acids)
- 4235 SEQ ID NO: 499: -0.282857, 106, a tail assembly protein, similar to tail assembly proteins for example ,GpK [Bacteriophage lambda] gi|139638|sp|P03729|VTAK#LAMB (97% identity in 199 amino acids)

SEQ ID NO: 500: -0.675172, 146, a tail assembly protein,
 4240 similar to tail assembly proteins for example ,GpI
 [Bacteriophage lambda] gi|139637|sp|P03730|VTAI#LAMBD
 (98% identity in 223 amino acids)

SEQ ID NO: 501: 0.114286, 64, a host specificity protein,
 similar to host specificity proteins for example ,GpJ
 4245 [Bacteriophage lambda] gi|138412|sp|P03749|VHSJ#LAMBD
 (89% identity in 1131 amino acids)

SEQ ID NO: 502: -0.550256, 196, a putative membrane
 protein precursor, similar to membrane protein Lom precursors
 for example ,lprophage P-EibA
 4250 gi|7532789|gb|aaF63231.1|AF151091#2 (69% identity in 199
 amino acids); [Bacteriophage lambda]
 gi|138693|sp|P03701|VLOM#LAMBD (44% identity in 199
 amino acids)

SEQ ID NO: 503: 0.15098, 52, a putative tail fiber protein,
 4255 similar to putative tail fiber proteins for example ,Gp37
 [Escherichia coli] gi|7466858|pir||G64887 (95% identity in
 496 amino acids)

SEQ ID NO: 504: 0.198571, 71, a tail fiber assembly protein,
 similar to tail fiber assembly proteins for example ,Orf194
 4260 [Bacteriophage lambda] gi|139990|sp|P03740|Y194#LAMBD
 (92% identity in 191 amino acids)

SEQ ID NO: 505: -0.96087, 93, novel, similar to hypothetical
 proteins for example ,putative catalase [Salmonella
 typhimurium] gi|7162108|emb|CAB76676.1| (84% identity in
 4265 289 amino acids)

SEQ ID NO: 506: -0.407736, 350, novel, similar to
 hypothetical proteins for example ,YciE [Escherichia coli]
 gi|775201|gb|aaA65179.1| (88% identity in 168 amino acids)

SEQ ID NO: 507: -0.273387, 125, novel, similar to
 4270 hypothetical proteins for example ,YciF [Escherichia coli]
 gi|140432|sp|P21362|YCIF#ECOLI (80% identity in 166 amino
 acids)

SEQ ID NO: 508 : -0.473626, 274, novel, similar to
hypothetical proteins for example ,YciG-homolog [Salmonella
4275 typhimurium] gi|6851081|emb|CAB71036.1| (88% identity in
60 amino acids), (also similar to YciG, E. coli [in TONB-TRPA
INTERGENIC REGION])

SEQ ID NO: 509: 0.544262, 62, novel, similar to hypothetical
proteins for example ,ybcY [Escherichia coli]
4280 gi|2495559|sp|P77460|YBCY#ECOLI (99% identity in 143
amino acids)

SEQ ID NO: 510 : -0.353615, 167, novel, similar to
hypothetical proteins for example ,YlcE [Escherichia coli]
gi|3025212|sp|P77087|YLCE#ECOLI (98% identity in 61
4285 amino acids), (similar to orf194, lambda, phage tail assembly
protein)

SEQ ID NO: 511 : -0.336744, 646, novel, similar to
hypothetical proteins for example ,L0013 [Escherichia coli
O-157:H7 EDL933] gi|3414881|gb|aaC31492.1| (99% identity
4290 in 133 amino acids)

SEQ ID NO: 512: 0.348333, 61, novel, similar to hypothetical
proteins for example ,L0014 [Escherichia coli O-157:H7
EDL933] gi|3414882|gb|aaC31493.1| (100% identity in 115
amino acids)

4295 SEQ ID NO: 513 : -0.398876, 90, novel, similar to
hypothetical proteins for example ,L0015 [Escherichia coli
O-157:H7 EDL933] gi|3414883|gb|aaC31494.1| (100% identity
in 512 amino acids)

SEQ ID NO: 514: 0.087324, 72, a putative fimbrial protein
4300 (partial), similar to truncated BfpA [Escherichia coli]
gi|4808944|gb|aaD30026.1|AF119170#1 (75% identity in 40
amino acids)

SEQ ID NO: 515 : -0.027193, 115, novel, similar to
hypothetical proteins for example ,[plasmid F]
4305 gi|8918853|dbj|Baa97900.1| (76% identity in 492 amino acids)

SEQ ID NO: 516: -0.440678, 178, an outer membrane protease

precursor, similar to outer membrane protease precursors for example ,protease VII precursor [Escherichia coli] gi|129161|sp|P09169|OMPT#ECOLI (98% identity in 317 amino acids)

4310 SEQ ID NO: 517 : -0.283069, 190, novel, similar to hypothetical proteins for example ,putative DNAbinding protein [Streptomyces coelicolor A3(2)] gi|6855358|emb|CAB71249.1| (34% identity in 171 amino acids)

4315 SEQ ID NO: 518: -0.234839, 156, a transposase, identical to hypothetical protein[Escherichia coli plasmid p O-157 insertion sequence IS629] gi|7444868|pir||T00241

4320 SEQ ID NO: 519: 0.076471, 69, a transposase, identical to transposase [Escherichia coli plasmid p O-157 insertion sequence IS629] gi|7443862|pir||T00240

SEQ ID NO: 520: 0.045946, 75, similar to a part of hypothetical proteins, for example, YPJA#ECOLI gi|2507221|sp|P52143 (amino acids at the position 1336-1569/1569) (96% identity in 234 amino acids), GTG start

4325 SEQ ID NO: 521: -0.288889, 73, novel

SEQ ID NO: 522 : 1.11087, 47, a transposase (insertion sequence IS629), similar to gi|7443862|pir|T00240 (96% identity in 296 amino acids)

4330 SEQ ID NO: 523 : -0.714754, 62, a transposase (insertion sequence IS629), similar to hypothetical proteins for example ,[Shigella flexneri SHI-2 pathogenicityisland] gi|5532454|gb|aaD44738.1|AF141323#9 (98% identity in 108 amino acids)

4335 SEQ ID NO: 524: -0.468595, 122, a putative TonB dependent outer membrane receptor, similar to TonBdependent outer membrane receptor PrrA [Escherichia coli CFT073] gi|3661477|gb|aaC61709.1| (97% identity in 656 amino acids)

SEQ ID NO:525: -0.648128, 188, a molybdenum transporter protein, similar to molybdenum transporter proteins for

4340

example ,gi|3661478|gb|aaC61710.1| (91% identity in 284 amino acids)

SEQ ID NO: 526: -0.117179, 554, novel

4345 SEQ ID NO: 527 : -0.148992, 646, novel, similar to hypothetical proteins for example ,Orf2 [Escherichia coli CFT073] gi|3661479|gb|aaC61711.1| (98% identity in 214 amino acids)

SEQ ID NO: 528 : -0.435414, 834, a putative ferric enterobactin transporter, similar to ferric

4350 enterobactin transporter ATP-binding protein [Escherichia coli CFT073] gi|3661480|gb|aaC61712.1| (79% identity in 148 amino acids)

SEQ ID NO: 529: -0.008333, 109, a putative ABC protein (permease), similar to ABC transporter permeases for

4355 example ,[Haemophilus influenzae] gi|2501391|sp|Q57130|YE71#HAEIN (40% identity in 323 amino acids)

SEQ ID NO: 530: -0.180172, 117, a putative ABC transporter, similar to iron (iii) ABC transporter, ATP-binding protein

4360 [Pyrococcus abyssi (strain Orsay)] gi|7519847|pir||A75077 (24% identity in 246 amino acids); hypothetical proteins for example ,[Methanosarcina barker] gi|2129363|pir||S62196 (26% identity in 259 amino acids)

SEQ ID NO: 531: -0.46554, 149, novel

4365 SEQ ID NO: 532: 0.172807, 115, a putative integrase, similar to phage integrase family, for example ,[Bacteriophage 21] gi|138558|sp|P27077|VINT#BPP21 (50% identity in 370 amino acids)

SEQ ID NO: 533: -0.333614, 239, a putative excisionase, similar to excisionases for example ,[Bacteriophage 21]

4370 gi|139674|sp|P27079|VXIS#BPP21 (45% identity in 77 amino acids)

SEQ ID NO: 534: -0.296774, 125, a putative exonuclease, its N-terminal part (amino acids at the position 1-256) is similar to

- 4375 hypothetical proteins for example ydfE [Escherichia coli
crypticphage/truncated insertion sequence IS2 fusion]
gi|78597|pir||S03698 (92% identity in 256 amino acids); its
Central part (amino acids at the position 209-622) is similar to
Exodeoxyribonuclease VIII (EC 3.1.11.-) (Exo VIII).
- 4380 [Escherichia coli] gi|1742216|dbj|Baa14950.1| (39% identity
in 361 amino acids); its C-terminal part (amino acids at the
position 644-776) is similar to exonuclease [phage T4]
gi|119690|sp|P04536|EXOD#BPT4 (27% identity in 133 amino
acids)
- 4385 SEQ ID NO: 535: -0.091398, 94, novel, similar to hypothetical
protein YdfD [Escherichia coli]
gi|140587|sp|P29010|YDFD#ECOLI (96% identity in 63 amino
acids)
- SEQ ID NO: 536: -0.238298, 142, a putative cell division
inhibition protein, similar to cell division inhibitor dicB
- 4390 [Escherichia coli] gi|2507009|sp|P09557|DICB#ECOLI (93%
identity in 62 amino acids)
- SEQ ID NO: 537: -0.317647, 953, novel
- SEQ ID NO: 538: -0.665487, 114, novel
- 4395 SEQ ID NO: 539: -0.364655, 233, novel, similar to
hypothetical 8.3 KD protein YdfC [Escherichia coli]
gi|140586|sp|P21418|YDFC#ECOLI, (94% identity in 72 amino
acids)
- SEQ ID NO: 540: -0.672619, 85, a putative repressor protein
- 4400 of division inhibition gene dicB, similar to DicA repressor
protein of division inhibition gene dicB [Escherichia coli]
gi|118631|sp|P06966|DICA#ECOLI (63% identity in 131 amino
acids); its N-terminal part (amino acids at the position 1-68
amino acids) is similar to N-terminal part of protein
- 4405 [Bacteriophage P22] gi|133359|sp|P03035|RPC2#BPP22(61%
identity in 68 amino acids)
- SEQ ID NO: 541: -0.47226, 293, a putative repressor protein
of division inhibition gene dicB, similar to DicC repressor

protein of division inhibition gene *dicB* [Escherichia coli
4410 gi|118633|sp|P06965|DICC#ECOLI (82% identity in 74 amino
acids); its N-terminal part (amino acids at the position 1-57
amino acids) is similar to (at low level) Cro [Bacteriophage P22]
gi|132195|sp|P09964|RCRO#BPP22 (36% identity in 57 amino
acids)

4415 SEQ ID NO: 542: -0.389388, 246, novel, similar to
hypothetical 11.0 kDa protein YdfX [Escherichia coli]
gi|3183265|sp|P76165|YDFX#ECOLI (87% identity in 93
amino acids)

4420 SEQ ID NO: 543: -0.211702, 95, novel, similar to replication
termination factor (prepriming protein D) DnaT [Escherichia
coli gi|1361001|pir|S56589 (51% identity in 83 amino acids)
SEQ ID NO: 544: -0.145524, 783, a putative phagereplication
protein, similar to phagereplication proteins for example ,
protein 14 [phage phi-80] gi|137937|sp|P14814|VG14#BPPH8
4425 (48% identity in 129 amino acids)

SEQ ID NO: 545: -0.473433, 1134, a putative fimbrial minor
pilin protein precursor, similar to N-terminal part of fimbrial
minor pilin protein precursors for example ,Pap-related pilus
H [Escherichia coli] gi|837337|gb|aaA67692.1| (75% identity in
4430 56 amino acids), GTG start, probably interrupted by frameshift
SEQ ID NO: 546: 0.168627, 52, a fimbrial minor pilin protein
precursor (partial), similar to C-terminal part of fimbrial minor
pilin protein precursors, for example ,PrsH [Escherichia coli]
gi|1172646|sp|P42185|PRSH#ECOLI (62% identity in 50
4435 amino acids)

SEQ ID NO: 547: 0.350336, 150, a putative colonization factor,
identical to Anm (attachment and effacement of negative
mutant) protein [Escherichia coli]
gi|6715555|gb|aaB48445.2| (100% identity in 252 amino
4440 acids); similar to accessory colonization factor AcfC [Vibrio
cholerae] gi|558481|gb|aaA50604.1| (50% identity in 239
amino acids)

Appendix B: Hideo *et al.* Full Translation

SEQ ID NO: 548: -0.544186, 302, a putative toxic protein (prophage maintenance; modulation host cell killing), similar to
4445 Hok/Gef family for example ,Gef [Escherichia coli] gi|2120017|pir|S40540 (79% identity in 69 amino acids)
SEQ ID NO: 549: -0.409434, 54, novel, similar to Rem protein [Escherichia coli] gi|132324|sp|P07010|REM#ECOL1 (71% identity in 84 amino acids)
4450 SEQ ID NO: 550: -0.517544, 58, novel, similar to (at low level) orf QD1 [Bacteriophage N15] gi|2564084|gb|aaB81659.1| (33% identity in 64 amino acids)
SEQ ID NO: 551: -0.641758, 92, novel, similar to hypothetical protein b1560 [Escherichia coli] gi|7466196|pir|C64911 (86%
4455 identity in 347 amino acids); similar to hypothetical protein A [phage P1] gi|732234|sp|Q06262|YORA#BPP1 (85% identity in 347 amino acids), GTG start
SEQ ID NO: 552: -0.407064, 454, a putative crossover junction endodeoxyribonuclease, similar to Gp67 [Bacteriophage HK97] gi|6901639|gb|aaF31142.1| (60% identity in 113 amino acids); crossover junction endodeoxyribonucleases for example ,Rus [Escherichia coli cryptic lambdoid prophage DLP12] gi|2507117|sp|P40116|RUS#ECOL1 (39% identity in 115 amino acids)
4460
4465 SEQ ID NO: 553: -0.475714, 71, novel
SEQ ID NO: 1213: -0.410758, 410, novel [hypothetical lipoprotein], its C-terminal part is similar to orf2 [Bacteriophage P27] gi|8346569|emb|CAB93762.1| (98% identity in 63 amino acids), GTG start
4470 SEQ ID NO: 1214: -0.622581, 63, a putative DNA methylase, similar to orf3 [BacteriophageP27] gi|8346570|emb|CAB93763.1| (85% identity in 312 amino acids); similar to adenine specific modification methylases for example ,Gp52 [phage N15] gi|7433503|pir|T13139 (55%
4475 identity in 270 amino acids)
SEQ ID NO: 1215: -0.359514, 248, novel, similar to

hypothetical proteins for example ,[Bacteriophage 933W]
gi|4585419|gb|aaD25447.1|AF125520#42 (52% identity in 613
amino acids)

4480 SEQ ID NO: 1216: 0.293846, 66, a putative holin protein,
similar to holin proteins for example ,[Bacteriophage 933W]
gi|4499808|emb|CAB39307.1| (95% identity in 71 amino acids)
SEQ ID NO: 1217 : 0.377049, 62, novel, similar to
hypothetical protein YdfR [Escherichia coli]
4485 gi|3183262|sp|P76160|YDFR#ECOLI (45% identity in 74
amino acids)
SEQ ID NO: 1218 : -0.180952, 64, a putative endolysin,
similar to endolysins for example ,[Bacteriophage 933W]
gi|4585422|gb|aaD25450.1|AF125520#45 (96% identity in 177
amino acids)

4490 SEQ ID NO: 1219 : -0.23625, 81, a putative antirepressor
protein, identical to putative antirepressor protein Ant
[Bacteriophage 933W] gi|4585423|gb|aaD25451.1|AF125520;
similar to antirepressor protein Ant [Bacteriophage
4495 P22|gi|131843|sp|P03037|RANT#BPP22 (49% identity in 126
amino acids)
SEQ ID NO: 1220 -0.936364, 100, endopeptidase (host lysis),
identical to Rz [Bacteriophage VT2-Sa]
gi|5881639|dbj|Baa84330.1|; similar to Rz endopeptidases for
example ,[Bacteriophage lambda]
4500 gi|119368|sp|P00726|ENPP#LAMBD (69% identity in 153
amino acids)
SEQ ID NO: 1221: -0.548598, 322, a lipoprotein Rz1 precursor,
similar to Rz1 protein precursors for
4505 example ,[Bacteriophage 933W]
gi|4585425|gb|aaD25453.1|AF125520#48(98% identity in 61
amino acids); [Bacteriophage lambda]
gi|540738|pir||JN0750(70% identity in 61 amino acids)
SEQ ID NO: 1222 : -0.179452, 74, novel, similar to
4510 hypothetical proteins for example ,[Escherichia coli]

gi|1778472|gb|aaB40755.1| (70% identity in 67 amino acids)
 SEQ ID NO: 1223: -0.636194, 269, a putative DNase, similar
 to (at low level) putative DNase [Bacteriophagephi-C31]
 gi|1107475|emb|Caa62587.1| (28% identity in 85 amino acids)
 4515 SEQ ID NO: 1224: 0.322807, 115, novel
 SEQ ID NO: 1225: -0.454217, 84, a putative terminase small
 subunit, similar to (at low level) putative terminase small
 subunit [Bacillus subtilis PBSX phage]
 gi|1722886|sp|P39785|XTMA#BACSU (42% identity in 57
 4520 amino acids), GTG start
 SEQ ID NO: 1226: -0.484559, 137, a putative terminase large
 subunit, similar to phage D3terminase-like protein
 [Haemophilus influenzae]
 gi|6739656|gb|aaF27357.1|AF198256#11 (22% identity in 472
 4525 amino acids)
 SEQ ID NO: 1227: -0.942222, 91, a putative head
 protein/prohead protease, its N-terminal part is similar to
 putative prohead proteases for example, [Bacteriophage
 HK97] gi|1722780|sp|P49860|VP4#BPHK7 (28% identity in
 4530 136 amino acids); its C-terminal part is similar to major head
 protein [mycobacterium phage L5]
 gi|465114|sp|Q05223|VG17#BPML5 (23% identity in 280
 amino acids), GTG start
 SEQ ID NO: 1228: -0.382433, 75, novel
 4535 SEQ ID NO: 1229: -0.597662, 386, a putative portal protein,
 its N-terminal-half part is similar to head portal proteins,
 for example, [Bacteriophage HK022]
 gi|6863114|gb|aaF30355.1|AF069308#3 (26% identity in 351
 amino acids); its C-terminal-half part is similar to
 4540 C-terminal-half part of putative transducer protein [H.
 salinarum] gi|3913878|sp|Q48317|HTR4#HALSA(21% identity
 in 347 amino acids)
 SEQ ID NO: 1230: -0.524865, 186, novel
 SEQ ID NO: 1231: -0.486352, 404, a putative head-tail

4545 adaptor, similar to putative head-tail adaptors for example ,[Bacteriophage HK97] gi|6901597|gb|aaF31100.1| (45% identity in 111 amino acids)
 SEQ ID NO: 1232: -0.194643, 113, novel, similar to phage hypothetical proteins for example ,Gp10 [Bacteriophage HK97]
 4550 gi|6901598|gb|aaF31101.1| (75% identity in 148 amino acids)
 SEQ ID NO: 1233: 0.009184, 99, novel, similar to Gp11 [Bacteriophage HK97] gi|6901599|gb|aaF31102.1| (49% identity in 113 amino acids)
 SEQ ID NO: 1234: -1.106849, 147, a putative major tail subunit, similar to major tail subunit [Bacteriophage HK97]
 4555 gi|6901588|gb|aaF31091.1|AF069529#4 (65% identity in 234 amino acids), GTG start
 SEQ ID NO: 1235: -1.563158, 58, a putative tail assembly chaperone, similar to putative tailassembly chaperons for example ,p14 [Bacteriophage HK97]
 4560 gi|6901600|gb|aaF31103.1| (62% identity in 124 amino acids)
 SEQ ID NO: 1236: -0.692373, 119, novel, similar to C-terminal part of Gp14 [Bacteriophage HK97]
 gi|6901601|gb|aaF31104.1|(60% identity in 94 amino acids),
 4565 probablyproduced by translational frameshift
 SEQ ID NO: 1237: -0.32604, 554, a putative tail length tape measure protein, similar to tail length tape measure proteins for example ,[Bacteriophage HK97]
 gi|6901589|gb|aaF31092.1|AF069529#5 (52% identity in 1022
 4570 amino acids)
 SEQ ID NO: 1238: -0.727957, 94, a putative minor tail protein, similar to minor tail proteins for example ,GpM [Bacteriophage lambda] gi|138845|sp|P03737|VMTM#LAMBD (44% identity in 110 amino acids), GTG start
 4575 SEQ ID NO: 1239: -0.284615, 92, a putative minor tail protein, similar to minor tail proteins for example ,GpL [Bacteriophage lambda] gi|138844|sp|P03738|VMTL#LAMBD (72% identity in 137 amino acids)

SEQ ID NO: 631: -0.709473, 381, a putative host specificity
 4580 protein, similar to host specificity proteins for example ,GpJ
 [Bacteriophage lambda] gi|138412|sp|P03749|VHSJ#LAMBD
 (69% identity in 1157 amino acids)
 SEQ ID NO: 632: -0.351282, 79, a putative outer membrane
 protein precursor, similar to outer membrane protein Lom
 4585 precursors for example ,[prophage P-EibA]
 gi|7532789|gb|aaF63231.1|AF151091#2(77% identity in 199
 amino acids); [Bacteriophage lambda]
 gi|138693|sp|P03701|VLOM#LAMBD (40% identity in 199
 amino acids)
 4590 SEQ ID NO: 633: -0.545985, 275, a putative tail fiber
 protein, similar to tail fiber proteins for
 example ,[Bacteriophage 933W]
 gi|4585436|gb|aaD25464.1|AF125520#59 (38% identity in 370
 amino acids)
 4595 SEQ ID NO: 634: -0.471244, 234, novel, similar to
 hypothetical protein [Bacteriophage 933W]
 gi|4585437|gb|aaD25465.1|AF125520#60 (92% identity in 89
 amino acids)
 SEQ ID NO: 635: -0.194, 101, novel
 4600 SEQ ID NO: 636: 1.042727, 111, novel, similar to hypothetical
 proteins for example ,Orf2 [Escherichia coli strain B171-8]
 gi|4126792|dbj|Baa36750.1| (37% identity in 111 amino acids)
 SEQ ID NO: 637: -0.138976, 509, novel
 SEQ ID NO: 638: -0.319205, 152, an integrase, similar to
 4605 integrases, for example, [Bacteriophage HK022]
 gi|138560|sp|P16407|VINT#BPHK0 (89% identity in 229
 amino acids), maybe comprising the deletion of 100 amino acids
 at N-terminus
 SEQ ID NO: 639: -0.625, 57, novel
 4610 SEQ ID NO: 640: -0.083333, 97, novel
 SEQ ID NO: 641: -0.538333, 121, disrupted transposase,
 similar to C-terminal of putative transposases for

example ,[Yersinia pestis plasmid pMT1]
 gi|2996347|gb|aaC13227.1|(74% identity in 89 amino acids),
 4615 TTG start
 SEQ ID NO: 642: -0.450655, 230, a disrupted transposase,
 similar to C-terminal part of putative transposases, for example,
 [Yersinia pestis plasmid pMT1] gi|7447905|pir|T14710 (70%
 identity in 90 amino acids), comprising the deletion of
 4620 N-terminal part (-180 amino acids)
 SEQ ID NO: 643: 0.76381, 106, novel, identical to L0015
 [Escherichia coli O-157:H7 strain EDL933]
 gi|3414883|gb|aaC31494.1|
 SEQ ID NO: 644: -0.675317, 159, novel, identical to L0014
 4625 [Escherichia coli O-157:H7 strain EDL933]
 gi|3288157|emb|Caa11510.1|
 SEQ ID NO: 645: -0.396079, 154, novel, identical to L0013
 [Escherichia coli O-157:H7 strain EDL933]
 gi|3414881|gb|aaC31492.1|
 4630 SEQ ID NO: 646: 0.016667, 61, novel
 SEQ ID NO: 647: 0.228866, 98, novel, similar to (at low level)
 hypothetical protein [insertion sequence IS630]
 gi|140943|spP16943|YIS5#SHISO (47% identity in 25 amino
 acids), TTG start
 4635 SEQ ID NO: 648: -0.455333, 151, novel
 SEQ ID NO: 649: -0.113235, 69, novel, similar to hypothetical
 proteins for example ,orf2 [Escherichia coli strain B171-8]
 gi|4126790|dbj|Baa36748.1|, (63% identity in 206 amino
 acids)
 4640 SEQ ID NO: 650: -1.015625, 65, bfpT-regulated chaperone-like
 protein, similar to TrcA (bfpT-r for example ,ulated
 chaperone-like protein)-like proteins for
 example ,TrcA[Escherichia coli strain B171-8]
 gi|4126789|dbj|Baa36747.1|, (72% identity in 195 amino
 4645 acids)
 SEQ ID NO: 651: -0.513812, 182, novel, partially similar to

hypothetical protein [insertion sequence IS630]
gi|140943|spP16943|YIS5#SHISO (81% identity in 60 amino
acids), GTG start, probably disrupted

4650 SEQ ID NO: 652: -0.585648, 642, novel, similar to N-terminal
part of hypothetical 39 kDa protein [insertion element IS630]
gi|1143207|gb|aaA84873.1| (82% identity in 54 amino acids)
SEQ ID NO: 653: -0.526471, 69, novel, similar to hypothetical
protein ORF2 [Escherichia coli strain B171-8]

4655 gi|4126790|dbj|Baa36748.1| (38% identity in 167 amino
acids); ORF4 [Escherichia coli strain B171-8]
gi|4126792|dbj|Baa36750.1| (40% identity in 127 amino acids)
SEQ ID NO: 654: -0.431519, 534, a putative transcription
regulatory protein, similar to transcription regulatory proteins

4660 for example ,UidR [Escherichia coli]
gi|2495429|sp|Q59431|UIDR#ECOLI (30% identity in 123
amino acids)
SEQ ID NO: 655 : -0.048747, 440, a putative
multidrug-effluxtransporter proteinprecursor, similar to

4665 multidrug-efflux transporter protein precursors for
example ,AcrA [Escherichia coli K-12]
gi|399000|sp|P31223|ACRA#ECOLI (51% identity in 358
amino acids)
SEQ ID NO: 656 : -0.159091, 111, a putative

4670 multidrug-effluxtransporter protein, similar to
multidrug-effluxtransporter proteins for example ,AcrB
[Escherichia coli K-12] gi|399001|sp|P31224|ACRB#ECOLI
(56% identity in 974 amino acids)
SEQ ID NO: 657: -0.38651, 342, a putative outer membrane

4675 channel protein, similar to outer membrane channel proteins
for example ,OprM [Pseudomonas aeruginosa]
gi|3184190|dbj|Baa28694.1| (43% identity in 448 amino acids)
SEQ ID NO: 658 : -0.231818, 133, a putative membrane
transporter protein, similar to membrane transporter protein

4680 for example ,[Streptomyces coelicolor A3(2)]

Appendix B: Hideo *et al.* Full Translation

gi|6469269|emb|CAB61730.1| (38% identity in 380 amino acids)

SEQ ID NO: 659 : -0.434188, 118, novel, similar to hypothetical protein [Xylella fastidiosa]

4685 gi|9106817|gb|aaF84556.1|AE003997#12 (38% identity in 209 amino acids)

SEQ ID NO: 660: -0.471354, 193, similar to C-terminal part of B1327#ECOLI gi|1787587(amino acids at the position 224-310/310) (33% identity in 87 amino acids)

4690 SEQ ID NO: 661: -0.156489, 132, similar to N-terminal part of B1327#ECOLI gi|1787587(amino acids at the position 22-123/310) (62% identity in 113amino acid)

SEQ ID NO: 662 : -0.247561, 247, a transposase (insertion sequence IS629), identical to gi|7443862|pir||T00240

4695 SEQ ID NO: 663 : -0.355, 141, a transposase (insertion sequence IS629), identical to gi|7444868|pir||T00241

SEQ ID NO: 664 : -0.182639, 145, a putative regulatory element, similar to(at low level) regulatory proteins for example ,regulatory protein CI (235 amino acids)

4700 [Bacteriophage HK022] gi|1350835|sp|P18680 (42% identity in 66 amino acids)

SEQ ID NO: 665 : -0.463487, 850, a putative regulatory element, similar to Cro [Bacteriophage HK022] gi|1350553|sp|P18679 (61% identity in 73 amino acids)

4705 SEQ ID NO: 666: -0.314679, 110, its C-terminal part (amino acids at the position 139-262 / 262) is similar to C-terminal part of YDAU#ECOLI gi|1787622 (amino acids at the position 162-285 / 285) (79% identity in 124 amino acids)

SEQ ID NO: 667: -0.4625, 233, novel

4710 SEQ ID NO: 668: -0.390688, 248, novel

SEQ ID NO: 669: 0.20583, 224, novel

SEQ ID NO: 670: -0.342491, 1133, novel

SEQ ID NO: 671: -0.326633, 200, novel, similar to N-terminal part of Eamino acid protein [Bacteriophage P22]

4715 gi|418207|sp|Q03544|VEaa#BPP22(88% identity in 42 amino acids)
 SEQ ID NO: 672: -0.27899, 972, novel, partially similar to hypothetical protein [Bacteriophage H19J]
 gi|4490348|emb|CAB38711.1| (70% identity in 54 amino acids); partially similar to part of Gp45 [Bacteriophage N15]
 4720 gi|7521552|pir|T13131 (57% identity in 47 amino acids)
 SEQ ID NO: 673: -0.308808, 194, possible methyltransferase, similar to methyltransferases for example ,cytosine-specific methyltransferase XorII [Xanthomonas oryzae pv.]
 4725 gi|1709171|sp|P52311|MTX2#XANOR (40% identity in 365 amino acids)
 SEQ ID NO: 674: -0.40473, 297, novel, similar to (at low level) hypothetical protein HI0983[Haemophilus influenzae]
 gi|1074592|pir|D64163 (26% identity in 138 amino acids)
 4730 SEQ ID NO: 675: -0.432143, 169, novel
 SEQ ID NO: 676: -0.448193, 167, novel, similar to Orf79 [Bacteriophage D3] gi|8895177|gb|aaF80835.1| (36% identity in 199 amino acids)
 SEQ ID NO: 677: -1.706667, 61, novel, similar to hypothetical proteins for example ,YbcO [Escherichia coli cryptic prophage DLP12] gi|7467043|pir|C64787 (57% identity in 96 amino acids); Gp66 [Bacteriophage HK97]
 gi|6901638|gb|aaF31141.1| (56% identity in 94 amino acids)
 SEQ ID NO: 678: -0.237063, 144, a putative aniterminator, similar to (at low level) antiterminator proteinQ [Bacteriophage 21] gi|4539484|emb|CAB39993.1| (22% identity in 168 amino acids)
 SEQ ID NO: 679: -0.446341, 83, novel, similar to putative TerB proteins for example ,[Deinococcus radiodurans]
 4745 gi|7473690|pir|C75302 (26% identity in 129 amino acids)
 SEQ ID NO: 680: -0.403175, 127, novel, GTG start
 SEQ ID NO: 681: 0.010435, 116, novel, similar to hypothetical proteins for example ,[Bacteriophage 933W]

gi|4585419|gb|aaD25447.1|AF125520#42 (53% identity in 613
4750 amino acids)
SEQ ID NO: 682: -0.445312, 513, a putative holin protein
(host cell lysis), similar to holin proteins for
example ,[Bacteriophage 933W] gi|4499808|emb|CAB39307.1|
(91% identity in 71 amino acids)
4755 SEQ ID NO: 683: -0.57037, 55, novel, similar to hypothetical
protein [Escherichia coli] gi|3183262|sp|P76160|YDFR#ECOLI
(43% identity in 74 amino acids)
SEQ ID NO: 684: -0.313158, 495, a endolysin (host cell lysis),
similar to endolysins for example ,[Bacteriophage 933W]
4760 gi|4335686|gb|aaD17382.1| (96% identity in 177 amino acids)
SEQ ID NO: 685: -0.652681, 318, a putative antirepressor,
identical to putative antirepressor [Bacteriophage 933W]
gi|4585423|gb|aaD25451.1|AF125520#46 (100% identity in
189 amino acids); its N-terminal part (amino acids at the
4765 position 1-126) is similar to antirepressor protein Ant
[Bacteriophage P22] gi|131843|sp|P03037|RANT#BPP22 (49%
identity in 126 amino acids)
SEQ ID NO: 686: -0.24433, 195, an endopeptidase (host cell
lysis), similar to endopeptidases for example ,[Bacteriophage
4770 VT2-Sal] gi|5881639|dbj|Baa84330.1| (100% identity in 155
amino acids)
SEQ ID NO: 687: -0.965741, 109, novel, similar to
hypothetical protein [Escherichia coli]
gi|1778472|gb|aaB40755.1| (70% identity in 67 amino acids);
4775 hypothetical protein [Salmonella dublin]
gi|3511132|gb|aaC33722.1| (70% identity in 49 amino acids)
SEQ ID NO: 688: -0.397973, 297, a putative DNase, similar to
(at low level) gp30 (DNase) [Bacteriophagephi-C31]
gi|1107475|emb|Caa62587.1| (28% identity in 85 amino acids);
4780 similar to (at low level) TerF-related protein [Deinococcus
radiodurans] gi|7473956|pir||C75599 (33% identity in 72
amino acids)

Appendix B: Hideo *et al.* Full Translation

SEQ ID NO: - : -0.413248, 235, novel

SEQ ID NO: - : 0.280303, 67, a putative terminase small
 4785 subunit, similar to phage terminase small subunits for
 example ,[Bacillus subtilis PBSX]
 gi|1722886|sp|P39785|XTMA#BACSU (34% identity in 52
 amino acids)

SEQ ID NO: 1641: -0.383784, 297, a putative terminase large
 4790 subunit, similar to phage hypothetical proteins, for
 example ,phage D3 terminase-like protein [Haemophilus
 influenzae] gi|6739656|gb|aaF27357.1|AF198256#11 (22%
 identity in 57 amino acids)

SEQ ID NO: 1642 : -0.942593, 109, a phage major head
 4795 protein/prohead protease, its C-terminal part is similar to
 major head proteins for example ,[Mycobacterium phageL5]
 gi|465114|sp|Q05223|VG17#BPML5 (22% identity in 306
 amino acids); its N-terminal part is similar to putative
 prohead proteases for example ,[Rhodobacter capsulatus]
 4800 gi|6467535|gb|aaF13181.1|AF181080#3 (30% identity in 133
 amino acids); similar to putative prohead protease
 [Rhodobacter capsulatus]
 gi|6467535|gb|aaF13181.1|AF181080#3 (30% identity in 133
 amino acids), GTG start

4805 SEQ ID NO: - : -0.615396, 657, novel

SEQ ID NO: 1419: 0.067253, 285, a putative portal protein,
 similar to phage portal proteins for example ,[Bacteriophage
 D3] gi|5059250|gb|aaD38955.1| (24% identity in 366 amino
 acids)

4810 SEQ ID NO: 1420: -0.121505, 94, novel

SEQ ID NO: 1421: -0.211215, 215, novel

SEQ ID NO: 1422: 0.150397, 253, a putative phage head-tail
 adaptor, similar to head-tail adaptors for
 example ,[Bacteriophage HK97] gi|6901597|gb|aaF31100.1|
 4815 (44% identity in 111 amino acids)

SEQ ID NO: 1423: 0.99049, 327, novel, similar to phage

hypothetical proteins for example ,Gp10 [Bacteriophage HK97]
gi|6901598|gb|aaF31101.1| (75% identity in 148 amino acids)
SEQ ID NO: 1424: -0.024118, 341, novel, similar to Gp11
4820 [Bacteriophage HK97] gi|6901599|gb|aaF31102.1| (49%
identity in 113 amino acids)
SEQ ID NO: 1425: 0.580303, 67, a major tail subunit, similar
to major tail subunit [Bacteriophage HK97]
gi|6901588|gb|aaF31091.1|AF069529#4 (67% identity in 234
4825 amino acids)
SEQ ID NO: 338: -0.622872, 377, a putative tail assembly
chaperon, similar to tail assembly chaperon Gp14
[Bacteriophage HKJ97] gi|6901600|gb|aaF31103.1| (62%
identity in 124 amino acids)
4830 SEQ ID NO: 339: -0.239024, 83, novel, similar to C-terminal
part of Gp14 [Bacteriophage HK97]
gi|6901601|gb|aaF31104.1|(60% identity in 94 amino
acids),probably produced by translational frameshift
SEQ ID NO: 340: -0.7548, 824, a putative tail length tape
4835 measure protein, similar to tail length tape measure
proteins for example ,[Bacteriophage HK97]
gi|6901589|gb|aaF31092.1|AF069529#5 (52% identity in 1022
amino acids)
SEQ ID NO: 341: 0.230159, 64, a putative tail component,
4840 similar to minor tail proteins for example ,GpM
[Bacteriophage lambda] gi|138845|sp|P03737|VMTM#LAMBD
(45% identity in 110 amino acids), GTG start
SEQ ID NO: 342: -0.180645, 63, a putative tail component,
similar to minor tail proteins for example ,GpL
4845 [Bacteriophage lambda] gi|138844|sp|P03738|VMTL#LAMBD
(75% identity in 232 amino acids)
SEQ ID NO: 343: -0.133766, 78, a putative tail assembly,
similar to tail assembly proteins for example ,GpK
[Bacteriophage lambda] gi|6901605|gb|aaF31108.1| (35%
4850 identity in 226 amino acids)

SEQ ID NO: 344: -0.166667, 136, a putative tail assembly, similar to tail assembly proteins for example ,GpI [Bacteriophage lambda] gi|139637|sp|P03730|VTAI#LAMBD (69% identity in 224 amino acids)

4855 SEQ ID NO: 345: -0.626389, 73, novel
SEQ ID NO: 346: -0.679259, 136, a putative superoxide dismutase, similar to copper/zinc-superoxide dismutases for example ,[Salmonella typhimurium] gi|2462699|emb|Caa73588.1| (58% identity in 175 amino acids)

4860 SEQ ID NO: 347: -0.498667, 76, a putative phage host specificity protein, similar to host specificity proteins for example ,GpJ [Bacteriophage lambda] gi|138412|sp|P03749|VHSJ#LAMBD (70% identity in 1164 amino acids)

4865 SEQ ID NO: 348: -0.345355, 184, similar to outer membrane proteins for example ,Lom protein [Bacteriophage P-EibA] dad|AF151091-2|aaF63231.1| (68% identity in 199 amino acids)

4870 SEQ ID NO: 349: -0.672832, 347, a putative tail fiber protein, similar to putative tail fiber proteins for example ,[Bacteriophage 933W] gi|4585436|gb|aaD25464.1|AF125520#59 (AF125520) (34% identity in 233 amino acids), GTG start

4875 SEQ ID NO: 350: -0.670588, 222, novel, similar to hypothetical protein [Bacteriophage 933W] gi|4585437|gb|aaD25465.1|AF125520#60 (94% identity in 129 amino acids), GTG start

SEQ ID NO: 351: -0.268932, 104, novel, similar to ORF4 [Escherichia coli strain B171-8] gi|4126792|dbj|Baa36750.1| (35% identity in 116 amino acids): ORF2 [Escherichia coli strain B171-8] gi|4126790|dbj|Baa36748.1| (28% identity in 171 amino acids)

4880 SEQ ID NO: 352: -0.120755, 54, novel, similar to ORF4

Appendix B: Hideo *et al.* Full Translation

4885 [Escherichia coli strain B171-8] gi|4126792|dbj|Baa36750.1|
(91% identity in 135 amino acids); ORF2 [Escherichia coli
strain B171-8] gi|4126790|dbj|Bamino acid36748.1| (43%
identity in 205 amino acids)
SEQ ID NO: 353: -0.368651, 253, novel, similar to ORF4

4890 [Escherichia coli B171-8] gi|4126792|dbj|Baa36750.1| (41%
identity in 135 amino acids); ORF2 [Escherichia coli strain
B171-8] gi|4126790|dbj|Baa 36748.1| (36% identity in 126
amino acids)
SEQ ID NO: 354: 0.292857, 71, similar to YDBL#ECOLI
4895 gi|1787648 (71% identity in 109 amino acids), but comprising
different N-terminal part and C-terminal part
SEQ ID NO: 355: 0.012941, 86, a putative ABC-type
transporter protein, similar to N-terminal part of ABC-type
transporter protein YdbA.2 [Escherichia coli]

4900 gi|7465766|pir||C48399 (amino acids at the position 1-1128 /
2020) (49% identity in 1011 amino acids)
SEQ ID NO: 356: -1.156522, 93, a putative ABC-type
transporter protein, similar to C-terminal part of ABC-type
transporter protein YdbA.2 [Escherichia coli]

4905 gi|7465766|pir||C48399(amino acids at the position
1220-2020/2020) (77% identity in 806 amino acids)
SEQ ID NO: 357: -0.396839, 349, novel
SEQ ID NO: 358: -0.287395, 120, novel
SEQ ID NO: 359: -0.428409, 177, novel

4910 SEQ ID NO: 360: 0.049057, 107, novel
SEQ ID NO: 361: -0.469602, 353, novel, similar to Vgr
proteins for example ,VgrE protein [Escherichia coli
gi|2920625|gb|aaC32465.1| (98% identity in 702 amino acids)
SEQ ID NO: 362: -0.206969, 618, a Rhs protein, similar to
4915 Rhs core proteins for example ,RhsD [Escherichia coli]
gi|1786706 (92% identity in 1281 amino acids) (Conserved in
E.coli K-12)
SEQ ID NO: 363: 0.095775, 72, novel, similar to (at low level)

IpaH protein IPA7#SHIFL gi|124813|sp|P18014 (35% identity
 4920 in 120 amino acids); YDDK#ECOLI gi|3183258|sp|P76123 (32%
 identity in 100 amino acids)
 SEQ ID NO: 364: -0.074561, 115, similar to outer membrane
 porin precursors for example ,NMPC#ECOLI gi|1786765 (67%
 identity in 343 amino acids), but comprising different
 4925 N-terminal part
 SEQ ID NO: 365: -0.466667, 178, novel, GTG start
 SEQ ID NO: 366: -0.283069, 190, a putative fimbrial
 chaperone protein precursor, similar to fimbrial chaperone
 protein precursors for example ,FocC [Escherichia coli]
 4930 gi|1169720|sp|P46008|FOCC#ECOLI (67% identity 206 amino
 acids)
 SEQ ID NO: 367: -0.472903, 156, a putative type 1 fimbrial
 protein precursor, similar to type 1 fimbrial protein precursors
 for example ,[Escherichia coli]
 4935 gi|729528|sp|P04128|FM1A#ECOLI (64% identity 186 amino
 acids)
 SEQ ID NO: 368: 0.214754, 62, novel, GTG start
 SEQ ID NO: 369: -0.717334, 76, a putative regulatory element,
 similar to araC-family transcription regulatory elementAdpA
 4940 [Streptomyces coelicolor A3(2)] gi|7544056|emb|CAB87229.1
 (41% identity in 316 amino acids)
 SEQ ID NO: 370: -0.468595, 122, a damage-inducible protein,
 similar to damage-inducible proteins for example ,DinI
 [Escherichia coli] gi|2498305|sp|Q47143|DINI#ECOLI (36%
 4945 identity in 72 amino acids)
 SEQ ID NO: 371: -1.029787, 48, novel, similar to hypothetical
 proteins for example ,ORF4 [Escherichia coli]
 gi|4126792|dbj|Baa36750.1| (43% identity in 131 amino
 acids); ORF2 [Escherichia coli] gi|4126790|dbj|Baa36748.1|
 4950 (35% identity in 126 amino acids)
 SEQ ID NO: 372: -0.648128, 188, novel, similar to
 hypothetical proteins for example ,ORF4 [Escherichia coli]

Appendix B: Hideo *et al.* Full Translation

gi|4126790|dbj|Baa36748.1| (43% identity in 206 amino acids); ORF2 [Escherichia coli gi|4126792|dbj|Baa36750.1|
4955 (91% identity in 135 amino acids)

SEQ ID NO: 373 : -0.117179, 554, novel, similar to hypothetical proteins for example ,ORF4 [Escherichia coli gi|4126792|dbj|Baa36750.1| (34% identity in 116 amino acids); ORF2 [Escherichia coli gi|4126790|dbj|Baa36748.1| (28% identity in 171 amino acids)]
4960

SEQ ID NO: 374 : -0.148992, 646, novel, similar to hypothetical protein[Bacteriophage 933W] gi|4585437|gb|aaD25465.1|AF125520#60 (93% identity in 89 amino acids)
4965

SEQ ID NO: 375: -0.831147, 62, a putative tail fiber protein, similar to C-terminal part of putative tail fiber protein [Bacteriophage 933W] gi|4585436|gb|aaD25464.1|AF125520#59 (100% identity in 92 amino acids), GTG start, probably disrupted
4970

SEQ ID NO: 376: -0.483469, 860, a putative tail fiber protein, similar to N-terminal part of tail fiber proteins for example ,Gp37 [Escherichia coli gi|7466858|pir|G64887(57% identity in 271 amino acids); orf-401 [Bacteriophage lambda] gi|140053|sp|P03764|Y401#LAMBDA (56% identity in 269 amino acids), probably interrupted
4975

SEQ ID NO: 377 : -0.061111, 109, a putative outer host membrane protein precursor, similar to Lom-like proteins for example ,[prophage P-EibA] gi|7532789|gb|aaF63231.1|AF151091#2 (68% identity in 199 amino acids); Lom [Bacteriophage lambda] gi|138693|sp|P03701|VLOM#LAMBDA (44% identity in 199 amino acids)
4980

SEQ ID NO: 378: -0.192241, 117, a phage tail protein (host specificity protein), similar to host specificity proteins for example ,GpJ [Bacteriophage
4985

lambda|gi|138412|sp|P03749|VHSJ#LAMBDA (65% identity in
 1158 amino acids)
 SEQ ID NO: 379: -0.512838, 149, a tail assembly protein,
 4990 similar to tail assembly proteins for example ,GpI
 [Bacteriophage lambda] gi|139637|sp|P03730|VTAI#LAMBDA
 (69% identity in 224 amino acids)
 SEQ ID NO: 380: 0.172807, 115, a tail assembly protein,
 similar to tail assembly proteins for example ,GpK
 4995 [Bacteriophage lambda] gi|139638|sp|P03729|VTAK#LAMBDA
 (85% identity in 174 amino acids)
 SEQ ID NO: 381: -0.337367, 282, a minor tail component,
 similar to minor tail proteins for example ,GpL
 [Bacteriophage lambda] gi|138844|sp|P03738|VMTL#LAMBDA
 5000 (76% identity in 232 amino acids)
 SEQ ID NO: 382: -0.296774, 125, a minor tail component,
 similar to minor tail proteins for example ,GpM
 [Bacteriophage lambda] gi|138845|sp|P03737|VMTM#LAMBDA
 (79% identity in 109 amino acids)
 5005 SEQ ID NO: 383: -0.091398, 94, a tail length determination,
 similar to tail length tape measure protein precursors for
 example ,GpH [Bacteriophage lambda]
 gi|138843|sp|P03736|VMTH#LAMBDA (51% identity in 870
 amino acids)
 5010 SEQ ID NO: 384: -0.319298, 1027, a minor tail component,
 similar to minor tail proteins for example ,GpG-T
 [Bacteriophage lambda] gi|7429179|pir|TLBPTL (67%
 identity in 134 amino acids), produced by translational
 frameshift
 5015 SEQ ID NO: 385: -0.624779, 114, a minor tail component,
 similar to minor tail proteins for example ,GpG [Bacteriophage
 lambda] gi|138842|sp|P03734|VMTG#LAMBDA (43% identity in
 140 amino acids)
 SEQ ID NO: 386: -0.477931, 146, novel, probably
 5020 corresponding to protein V [Bacteriophage lambda]

- SEQ ID NO: 387: -0.276079, 1159, a minor tail component, similar to minor tail protein GpU [Bacteriophage lambda] gi|138847|sp|P03732|VMTU#LAMBDB (80% identity in 131 amino acids)
- 5025 SEQ ID NO: 388: -0.29799, 200, a minor tail component, similar to minor tail proteins for example ,GpZ [Bacteriophage lambda] gi|138849|sp|P03731|VMTZ#LAMBDB (69% identity in 177 amino acids)
- SEQ ID NO: 389: -0.661327, 438, a tail attachment (minor capsid protein), similar to minor capsid proteins for example ,GpFII [Bacteriophage lambda] gi|137575|sp|P03714|VCF2#LAMBDB (91% identity in 117 amino acids)
- 5030 SEQ ID NO: 390: -0.392135, 90, DNA-packaging, similar to DNA-packaging proteins for example ,GpFI [Bacteriophage lambda] gi|139324|sp|P03709|VPF1#LAMBDB (98% identity in 132 amino acids)
- 5035 SEQ ID NO: 391: 0.522727, 89, a major capsid protein, similar to major capsid proteins for example ,GpE [Bacteriophage lambda] gi|116701|sp|P05481|HEAD#BPPH8 (87% identity in 341 amino acids)
- 5040 SEQ ID NO: 392: -0.269369, 112, a head decoration protein (major capsid protein), similar to major capsid proteins for example ,GpD [Bacteriophage lambda] gi|137566|sp|P03712|VCAD#LAMBDB (99% identity in 110 amino acids)
- 5045 SEQ ID NO: 393: -0.239229, 442, a minor capsid protein precursor, similar to minor capsid protein precursors for example ,GpC [Bacteriophage lambda] gi|137565|sp|P03711|VCAC#LAMBDB (97% identity in 439 amino acids), capsid assembly protein containing Nu3-homolog
- 5050 SEQ ID NO: 394: -0.247826, 231, a portal protein (minor capsid protein), similar to portal proteins for example ,GpB [Bacteriophage lambda] gi|138762|sp|P03710|VMCB#LAMBDB

- 5055 (98% identity in 533 amino acids)
 SEQ ID NO: 395: -0.441584, 304, a head-to-tail joining, similar to head-to-tail joining proteins for example ,GpW [Bacteriophage lambda] gi|138415|sp|P03727|VHTJ#LAMBD (98% identity in 68 amino acids)
- 5060 SEQ ID NO: 396: -0.434392, 190, a terminase large subunit (DNA-packaging protein), similar to terminase large subunits for example ,GpA [Bacteriophage lambda] gi|137616|sp|P03708|TERL#LAMBD (97% identity in 641 amino acids), GTG start
- 5065 SEQ ID NO: 397: -0.085882, 86, a putative terminase small subunit, similar to terminasesmall subunits for example ,Nu1 [Bacteriophage lambda] (82% identity in 180 amino acids)
 SEQ ID NO: 398: -0.327551, 99, novel, similar to hypothetical proteins for example ,[Escherichia coli]
- 5070 gi|1778472|gb|aaB40755.1| (90% identity in 53 amino acids)
 SEQ ID NO: 399: -0.445312, 513, a putative transcription regulatory element, similar to PerC (BfpW) transcription activator eaeA/bfpA [Escherichia coli] gi|1172431|sp|P43475|PERC#ECOLI (47% identity in 87 amino acids)
- 5075 SEQ ID NO: 400: 0.010435, 116, a putative lipoproteinRz1 precursor, similar to lipoproteinRz1 precursors for example ,[phage lambda] gi|540738|pir|JN0750 (70% identity in 61 amino acids)
- 5080 SEQ ID NO: 401: -0.403175, 127, a putative host cell lysis, similar to endopeptidases for example ,[Bacteriophage H-19B] gi|4335687|gb|aaD17383.1| (77% identity in 150 amino acids)
 SEQ ID NO: 402: -0.542391, 93, novel, partially similar to hypothetical protein YchG [Escherichia coli]
- 5085 gi|267475|sp|P30192|YCHG#ECOLI (80% identity in 30 amino acids)
 SEQ ID NO: 403: -0.42, 51, novel, partially similar to a hypothetical proteins for example ,YchG[Escherichia coli]

gi|267475|sp|P30192|YCHG#ECOLI (95% identity in 60 amino
5090 acids), GTG start
SEQ ID NO: 404: -0.364583, 49, novel, partially similar to a
hypothetical protein b1240 [Escherichia coli]
gi|7466155|pir|C64871 (54% identity in 51 amino acids)
[0020]

5095 3) Proteins comprising Insertion Sequence; IS
Sequence number: hydrophobicity. The number of amino
acids. Character such as function
SEQ ID NO: 405: -0.221861, 216, novel, identical to
hypothetical protein [Bacteriophage VT2-Sa]
5100 gi|5881622|dbj|Baa84313.1|, but [having] different start;
similar to hypothetical protein [Bacteriophage 933W]
gi|4499790|emb|CAB39289.1| (85% identity in 78 amino acids)
SEQ ID NO: 406: -0.313776, 197, novel, identical to
hypothetical protein [Bacteriophage VT2-Sa]
5105 gi|5881623|dbj|Baa84314.1|, but [having] different start;
similar to hypothetical proteins for example ,NinB protein
[Bacteriophage 21](43% identity in 147 amino acids)
SEQ ID NO: 407: -0.486667, 61, a putative DNA methylase,
identical to hypothetical protein [Bacteriophage VT2-Sa]
5110 gi|5881624|dbj|Baa84315.1| (100% identity in 175 amino
acids); similar to hypothetical protein Gp62 [Bacteriophage
HK97] gi|6901634|gb|aaF31137.1| (98% identity in 175 amino
acids); similar to (at low level) DNA
N⁶-adenine-methyltransferase (M.T1) [Enterobacteria phage
5115 T1] gi|166164|gb|aaA87390.1| (31% identity in 143 amino
acids)
SEQ ID NO: 408: -0.175926, 55, novel, identical to
hypothetical protein [Bacteriophage VT2-Sa]
gi|5881625|dbj|Baa84316.1| (100% identity in 60 amino
5120 acids); similar to hypothetical proteins for example ,NinE
protein [Bacteriophage 21]gi|4539480|emb|CAB39989.1| (98%
identity in 60 amino acids)

SEQ ID NO: 409 : -0.017752, 170, novel, identical to
hypothetical protein [Bacteriophage VT2-Sa]
5125 gi|5881626|dbj|Baa84317.1| (100% identity in 57 amino acids),
GTG start

SEQ ID NO: 1375 : -0.38883, 189, a putative antirepressor
protein, identical to hypothetical protein [Bacteriophage
VT2-Sa] gi|5881627|dbj|Baa84318.1| (100% identity in 244
5130 amino acids); its C-terminal part similar to C-terminal part
antirepressor protein Ant [Bacteriophage P22]
gi|131843|sp|P03037|RANT#BPP22 (82% identity in 104
amino acids), its N-terminal part similar to N-terminal part of
hypothetical protein [Bacteriophage TP901-1]
5135 gi|2924237|emb|Caa74615.1| (42% identity in 119 amino
acids)

SEQ ID NO: 1376 : -0.209115, 374, a DNA-binding protein,
identical to Roi [Bacteriophage VT2-Sa]
gi|5881628|dbj|Baa84319.1|, but [having] different start;
5140 similar to Roi proteins for example ,[Enterobacteria phage
HK022] gi|1197729|gb|aa C48863.1| (82% identity in 242
amino acids)

SEQ ID NO: 1377 : 0.177508, 1028, novel, identical to
hypothetical protein orf15[Bacteriophage 933W]
5145 gi|4499798|emb|CAB39297.1| (100% identity in 201 amino
acids), similar to hypothetical proteins for example ,NinG
protein [Bacteriophage 21] gi|4539482|emb|CAB39991.1| (94%
identity in 201 amino acids)

SEQ ID NO: 1378 : -0.144201, 458, novel, identical to
5150 hypothetical protein orf16[Bacteriophage 933W]
gi|4499799|emb|CAB39298.1|(100% identity in 64 amino
acids); similar to hypothetical proteins for example ,Nin68
[Bacteriophage lambda]gi|9626304|ref|NP#040640.1|(80%
identity in 60 amino acids)

5155 SEQ ID NO: 1379: 0.890181, 388, antitermination protein Q,
identical to antitermination Q protein [Bacteriophage 933W]

gi|4585416|gb|aaD25444.1|AF125520#39, but [having]
different start; similar to antitermination Q proteins for
example ,[Bacteriophage H-19B] gi|2668768|gb|aaD04655.1|
5160 (96% identity in 144 amino acids)
SEQ ID NO: - : -0.090909, 221, novel, partially similar to
hypothetical protein [Bacteriophage P27]
gi|8346570|emb|CAB93763.1| (89% identity in 37 amino
acids), TTG start

5165 SEQ ID NO: 1676: 0.087912, 92, a Shiga toxin 2 subunit A,
identical to gi|1351074|sp|P09385|SLTA#BP933; identical to
ECs1908: Comp.(1899924-1900292), -0.25, 123, Shiga toxin 2
subunit B gi|134538|sp|P09386|SLTB#BP933
SEQ ID NO: 1644 : -0.397973, 297, novel, identical to
5170 N-terminal part of hypothetical protein [Bacteriophage 933W]
gi|4585419|gb|aaD25447.1|AF125520#42 (100% identity in
557 amino acids) ; similar to N-terminal part of hypothetical
proteins for example ,Yjhs [Shigella dysenteriae]
gi|6759965|gb|aaF28123.1|AF153317#19 (78% identity in 554
5175 amino acids)
SEQ ID NO: 1645 : -0.965741, 109, a transposase (OrfB)
(insertion sequenceIS629), identical to
gi|7443862|pir|T00240
SEQ ID NO: 1681 : -0.893204, 104, a transposase (OrfA)
5180 (insertion sequenceIS629), identical to
gi|7444868|pir|T00241 (100% identity in 108 amino acids)
SEQ ID NO: - : -0.342857, 85, novel, identical to hypothetical
protein [Bacteriophage 933W] gi|4499806|emb|CAB39305.1|
(100% identity in 59 amino acids)

5185 SEQ ID NO: - : -0.577099, 263, novel, identical to
hypothetical protein [Bacteriophage 933W]
gi|4585420|gb|aaD25448.1|AF125520#43 (100% identity in
148 amino acids)
SEQ ID NO: 877: -0.830769, 79, a putative holin protein,
5190 identical to protein [Bacteriophage VT2-Sa]

gi|5881636|dbj|Baa84327.1|; similar to putative holin proteins for example ,[Shigella dysenteriae] gi|6759967|gb|aaF28125.1|AF153317#21 (95% identity in 71 amino acids)

5195 SEQ ID NO: 878: -0.5, 141, a endolysin, identical to putative endolysin [Bacteriophage 933W] gi|4585422|gb|aaD25450.1|AF125520#45 (100% identity in 177 amino acids); similar to putative endolysins for example ,[Bacteriophage H-19B]

5200 gi|4335686|gb|aaD17382.1|(93% identity in 177 amino acids) SEQ ID NO: 879: -1.08, 71, a putative antirepressor protein, similar to identical to putative antirepressor protein [Bacteriophage 933W] gi|4585423|gb|aaD25451.1|AF125520#46; antirepressor proteinAnt [Bacteriophage P22] gi|131843|sp|P03037|RANT#BPP22 (49% identity in 121 amino acids)

5205 SEQ ID NO: 880: -0.375862, 88, a putative endopeptidase, identical to endopeptidase[Bacteriophage 933W]

5210 gi|4585424|gb|aaD25452.1|AF125520#47 (100% identity in 154 amino acids); similar to endopeptidases for example ,Rz [Bacteriophage lambda] gi|119368|sp|P00726|ENPP#LAMBD (72% identity in 154 amino acids)

5215 SEQ ID NO: 881: -0.477359, 54, a putative lipoproteinRz1 precursor, identical to putative Rz1 protein precursor [Bacteriophage 933W] gi|4585425|gb|aaD25453.1|AF125520#48(100% identity in 61 amino acids); similar to lipoproteinRz1 precursor [Bacteriophage lambda] gi|1017781|gb|aaC48862.1| (72% identity in 61 amino acids)

5220 SEQ ID NO: 882: -0.293827, 82, a Bor protein precursor, identical to [Bacteriophage 933W] gi|4585426|gb|aaD25454.1|AF125520#49 (100% identity in 97 amino acids); similar to Bor protein precursor [Bacteriophage

Appendix B: Hideo *et al.* Full Translation

5225 lambda|gi|137520|sp|P26814|VBOR#LAMBD (96% identity in
97 amino acids)
SEQ ID NO: 883 : -0.305483, 384, novel, similar to
hypothetical protein [Bacteriophage VT2-Sa|
gi|5881640|dbj|Baa84331.1| (85% identity in 75 amino acids)
5230 SEQ ID NO: 884: -0.434955, 330, a putative small subunit
terminase, identical to putative small subunit terminase
[Bacteriophage 933W]
gi|4585427|gb|aaD25455.1|AF125520#50 (100% identity in
268 amino acids)
5235 SEQ ID NO: 885: -0.576025, 464, a putative terminase large
subunit, identical to putative terminase large subunit
[Bacteriophage 933W]
gi|4585428|gb|aaD25456.1|AF125520#51 (100% identity in
568 amino acids)
5240 SEQ ID NO: 886: -0.238694, 200, a putative portal protein,
identical to putative portal protein [Bacteriophage 933W]
gi|4585429|gb|aaD25457.1|AF125520#52 (100% identity in
714 amino acids)
SEQ ID NO: 887 : -0.438542, 97, novel, identical to
5245 hypothetical protein [Bacteriophage 933W]
gi|4585430|gb|aaD25458.1|AF125520#53 (100% identity in
335 amino acids)
SEQ ID NO: 888 : -0.264131, 185, novel, identical to
hypothetical protein [Bacteriophage 933W]
5250 gi|4585431|gb|aaD25459.1|AF125520#54 (100% identity in
404 amino acids)
SEQ ID NO: 889 : -0.237063, 144, novel, identical to
hypothetical protein [Bacteriophage 933W]
gi|4585432|gb|aaD25460.1|AF125520#55 (100% identity in
5255 129 amino acids)
SEQ ID NO: 890 : 1.472727, 56, novel, identical to
hypothetical protein [Bacteriophage 933W]
gi|4585433|gb|aaD25461.1|AF125520#56 but [having]

different start

5260 SEQ ID NO: 891 : -0.255915, 618, novel, identical to
hypothetical protein [Bacteriophage 933W]
gi|4585434|gb|aaD25462.1|AF125520#57, but [having]
different start

5265 SEQ ID NO: 892 : 0.052113, 72, novel, identical to
hypothetical protein [Bacteriophage 933W]
gi|4585435|gb|aaD25463.1|AF125520#58 (100% identity in
216 amino acids)

5270 SEQ ID NO: 893: -0.046491, 115, a putative tail fiber protein,
identical to putative tail fiber protein [Bacteriophage 933W]
gi|4585436|gb|aaD25464.1|AF125520#59(100% identity in 645
amino acids)

5275 SEQ ID NO: 894 : -0.466667, 178, novel, identical to
hypothetical protein [Bacteriophage 933W]
gi|4585437|gb|aaD25465.1|AF125520#60, but [having]
different start

5280 SEQ ID NO: 895 : -0.283069, 190, novel, identical to
hypothetical protein [Bacteriophage 933W]
gi|4585438|gb|aaD25466.1|AF125520#61, but [having]
different start

5285 SEQ ID NO: 896 : -0.472903, 156, novel [putative outer
membrane protein; OMP], TTG start

5290 SEQ ID NO: 897: -0.717334, 76, novel [periplasmic], identical
to hypothetical protein [Bacteriophage 933W]
gi|4585439|gb|aaD25467.1|AF125520#62 (100% identity in
567 amino acids) ; its N-terminal part similar to hypothetical
protein [Bacteriophage P-EibD]
gi|7523538|gb|aaF63043.1|AF151675#5 (98% identity in 147
amino acids), GTG start

5290 SEQ ID NO: 898: -0.468595, 122, a putative tail tip fiber
protein, identical to hypothetical protein [Bacteriophage
933W] gi|4585440|gb|aaD25468.1|AF125520#63 (100%
identity in 422 amino acids); similar to(at low level) tail tip

- fiber protein gp21 [phage N15] gi|7444604|pir||T13107 (24% identity in 381 amino acids)
- 5295 SEQ ID NO: 899: -1.029787, 48, novel [putative outer membrane protein: OMP], identical to hypothetical protein [Bacteriophage 933W] gi|4585441|gb|aaD25469.1|AF125520#64, but [having] different start, TTG start
- 5300 SEQ ID NO: 900: -0.648128, 188, novel [putative outer membrane protein: OMP], identical to hypothetical protein [Bacteriophage 933W] gi|4585442|gb|aaD25470.1|AF125520#65 (100% identity in 205 amino acids)
- 5305 SEQ ID NO: 901: -0.117179, 554, a putative outer membrane precursor, identical to putative Lom precursor [Bacteriophage 933W] gi|4585443|gb|aaD25471.1|AF125520#66 (100% identity in 244 amino acids); similar to outer membrane proteinrck [Salmonella typhimurium] gi|282013|pir||A43309
- 5310 (35% identity in 172 amino acids); outer membrane protein Lom precursor gi|138693|sp|P03701|VLOM#LAMBDA (35% identity in 167 amino acids); ail gene products for example ,[Yersinia pseudotuberculosis] gi|5902750|sp|Q56957|AIL#YERPS (32% identity in 241 amino acids); virulence proteinpagC precursor [Salmonella typhimurium]gi|129558|sp|P23988|PAGC#SALTY (29% identity in 180 amino acids)
- SEQ ID NO: 902: -0.148992, 646, novel, identical to hypothetical protein [Bacteriophage 933W] gi|4585444|gb|aaD25472.1|AF125520#67 (100% identity in 133 amino acids)
- 5320 SEQ ID NO: 903: -0.831147, 62, novel, similar to hypothetical protein [Bacteriophage 933W] gi|4585445|gb|aaD25473.1|AF125520#68 (100% identity in 218 amino acids)
- 5325 SEQ ID NO: 904: -0.482819, 455, novel, identical to

Appendix B: Hideo *et al.* Full Translation

hypothetical protein [Bacteriophage 933W]
gi|4585446|gb|aaD25474.1|AF125520#69 (100% identity in
148 amino acids)

5330 SEQ ID NO: 905 : -0.420639, 408, novel, identical to
hypothetical protein [Bacteriophage 933W]
gi|4585447|gb|aaD25475.1|AF125520#70 (100% identity in 83
amino acids)

5335 SEQ ID NO: 906 : -0.063889, 109, novel, identical to
hypothetical protein [Bacteriophage 933W]
gi|4585448|gb|aaD25476.1|AF125520#71 (100% identity in
421 amino acids)

5340 SEQ ID NO: 907 : -0.171552, 117, novel, similar to
hypothetical protein [Bacteriophage 933W]
gi|4585449|gb|aaD25477.1|AF125520#72 (99% identity in
2793 amino acids)

5345 SEQ ID NO: 908 : -0.512838, 149, novel, identical to
hypothetical protein [Bacteriophage 933W]
gi|4585450|gb|aaD25478.1|AF125520#73, but [having]
different start

5350 SEQ ID NO: 909 : 0.189474, 115, novel, identical to
hypothetical protein [Bacteriophage 933W]
gi|9632540|ref|NP#049534.1| (100% identity in 114 amino
acids); similar to hypothetical proteins for example ,ygiW
protein precursor [Escherichia coli]
gi|1723887|sp|P52083|YGIW#ECOLI (53% identity in 93
amino acids)

5355 SEQ ID NO: 910: -0.313446, 239, a MokW protein (prophage
maintenance/modulation of host cell killing), identical to MokW
[Bacteriophage 933W]
gi|4585453|gb|aaD25481.1|AF125520#76 (100% identity in 70
amino acids); similar to GelF [Escherichia coli]
gi|1786200|gb|aaC73129.1| (73% identity in 69 amino acids)

5360 SEQ ID NO: 911 : -0.276613, 125, novel, identical to
hypothetical protein [Bacteriophage 933W]

gi|4585454|gb|aaD25482.1|AF125520#77, but [having]
different start

SEQ ID NO: 912 : -0.091398, 94, novel, identical to
hypothetical protein [Bacteriophage VT2-Sa]

5365 gi|5881668|dbj|Baa84359.1| (100% identity in 219 amino
acids); identical to C-terminal part of hypothetical protein
[Bacteriophage 933W]
gi|4585455|gb|aaD25483.1|AF125520#78(100% identity in 219
amino acids)

5370 SEQ ID NO: 913 : -0.343275, 1027, novel, identical to
hypothetical protein [Bacteriophage VT2-Sa]
gi|5881669|dbj|Baa84360.1|, but [having] different start;
similar to hypothetical protein [Bacteriophage 933W]
gi|7649907|dbj|Baa94185.1| (92% identity in 72 amino acids);

5375 hypothetical proteins for example ,[Bacteriophage VT2-Sa]
gi|4585386|gb|aaD25414.1|AF125520#9 (92% identity in 68
amino acids)

SEQ ID NO: 914 : -0.624779, 114, novel, identical to
hypothetical protein [Bacteriophage VT2-Sa]

5380 gi|5881670|dbj|Baa84361.1| (100% identity in 94 amino acids),
GTG start

SEQ ID NO: 915 : -0.332759, 233, novel, identical to
hypothetical protein [Bacteriophage VT2-Sa]
gi|5881671|dbj|Baa84362.1| (100% identity in 73 amino
acids); similar to C4-type zinc finger proteins (TraR family)
for example ,orf39 [Pseudomonas aeruginosa phage phi CTX]
gi|4063813|dbj|Baa36267.1| (42% identity in 59 amino acids)

5385 SEQ ID NO: 916: -0.407287, 248, a putative anti-repressor
protein, identical to hypothetical protein [Bacteriophage
VT2-Sa] gi|5881672|dbj|Baa84363.1|(100% identity in 209
amino acids); similar to hypothetical protein HI1422
[Haemophilus influenzae Rd]
gi|1175795|sp|P44193|YE22#HAEIN (40% identity in 158
amino acids); putative phage anti-repressor proteins for

- 5395 example ,[Neisseria meningitidis]
gi|7379969|emb|CAB84545.1| (49% identity in 112 amino acids)
SEQ ID NO: 917: 0.069027, 227, novel
SEQ ID NO: 918: -1.014706, 69, probably resistance to phage
- 5400 N4, lambda, Rtn membrane associated protein [Escherichia coli]
gi|2498867|sp|P76446|RTN#ECOLI (31% identity in 498 amino acids)
SEQ ID NO: 919: -0.130857, 176, novel, similar to FidL-Salmonella typhimurium
gi|4324611|gb|aaD16955.1| (29% identity in 149 amino acids)
- 5405 SEQ ID NO: 920 : -0.304721, 1166, a putative transcriptionactivator, similar to transcriptionactivators for example ,MarT [Salmonella typhimurium]
gi|4324612|gb|aaD16956.1| (28% identity in 268 amino acids)
- 5410 SEQ ID NO: 921: -0.308543, 200, a putative oxidoreductase, similar to oxidoreductases for example ,[Escherichia coli]
gi|2492762|sp|P76633|YGCW#ECOLI (55% identity in 257 amino acids)
SEQ ID NO: 922 : -0.814127, 362, a putative chaperone, similar to hypothetical proteins for example ,ORF60 [Yersinia pestis]
gi|7467334|pir||T17432 (48% identity in 204 amino acids); chaperone proteins for example ,EcpD [Escherichia coli]
gi|2506408|sp|P33128|ECPD#ECOLI (35% identity in 185 amino acids)
- 5415 SEQ ID NO: 923 : -0.431859, 114, novel, similar to hypothetical proteins for example ,ORF59 [Yersinia pestis]
gi|4106627|emb|Caa21382.1| (34% identity in 438 amino acids)
SEQ ID NO: 924: -0.114136, 192, a putative outer membrane usher protein, similar to hypothetical protein ORF 58 [Yersinia pestis]
gi|4106626|emb|Caa21381.1| (44% identity in 824 amino acids); outer membrane usher proteins for example ,FimD [Salmonella typhimurium]

gi|585135|sp|P37924|FIMD#SALTY (32% identity in 832
5430 amino acids)
SEQ ID NO: 925: -0.282297, 210, a putative chaperone,
similar to hypothetical protein ORF57 [Yersinia pestis]
gi|4106625|emb|Caa21380.1| (39% identity in 233 amino
acids); chaperone proteins for example ,EcpD [Escherichia
5435 coli] gi|2506408|sp|P33128|ECPD#ECOLI (36% identity in 217
amino acids)
SEQ ID NO: 926: -0.123005, 214, a putative pilin protein,
similar to hypothetical protein ORF56 [Yersinia pestis]
gi|4106624|emb|Caa21379.1| (36% identity in 185 amino
5440 acids); major pilin proteins for example ,Sf amino acids
[Escherichia coli] gi|4105989|gb|aaD02646.1| (32% identity in
181 amino acids)
SEQ ID NO: - : -0.309259, 109, novel
SEQ ID NO: 1488: -0.323145, 1012, a putative filamentous
5445 hemagglutinin-like protein, similar to
hemagglutinin/hemolysin-related proteins [Neisseria
meningitidis] for example ,gi|7225719|gb|aaF40927.1| (25%
identity in 1001 amino acids); filamentous hemagglutinin B
precursor [Bordetella pertussis] gi|78213|pir|S21010(20%
5450 identity in 824 amino acids)
SEQ ID NO: - : -0.353779, 808, a putative hemolysin
activatorrelated protein, similar to hemolysin activatorrelated
proteins for example ,[Pectobacterium chrysanthemi]
gi|1772622|gb|aaC31980.1| (27% identity in 484 amino
5455 acids);hemolysin activation protein precursor [Serratia
marcescens] gi|123205|sp|P15321|HLYB#SERMA (24%
identity in 475 amino acids)
SEQ ID NO: 1608 : -0.270213, 142, a putative
holo-[acyl-carrier protein] synthase, similar to
5460 holo-[acyl-carrier protein] synthases for
example ,[Campylobacter jejuni] gi|6968838|emb|CAB73833.1|
(39% identity in 121 amino acids)

SEQ ID NO: 1609: -0.224107, 113, a putative 3-oxoacyl-(acyl carrier protein) reductase, similar to 3-oxoacyl-(acyl carrier protein) reductases for example ,[Moritella marinal
5465 gi|7227179|gb|aaF42251.1| (41% identity in 188 amino acids)
SEQ ID NO: 1610 : -0.570629, 144, a putative (3R)-hydroxymyristol-(acyl carrier protein) dehydratase, similar to (3R)-hydroxymyristol-(acyl carrier protein)
5470 dehydratases for example ,gi|7190847|gb|aaF39621.1 (30% identity in 158 amino acids)
SEQ ID NO: 1611: -0.0544, 126, a putative acyl carrier protein, similar to acyl carrier proteins for example ,AcpC [Streptococcus agalactiae]
5475 gi|4886773|gb|aaD32036.1|AF093787#4 (38% identity in 86 amino acids)
SEQ ID NO: 1409: -0.480057, 703, a putative aminomethyl transferase, similar to aminomethyl transferases for example ,gi|7450600|pir||C75088 (26% identity in 333 amino
5480 acids)
SEQ ID NO: 1410 : -0.678001, 1401, a putative 3-oxoacyl-[acyl-carrier- protein] synthase, its N-terminal-half part is similar to 3-oxoacyl-[acyl-carrier- protein] synthase (EC 2.3.1.41) [Bacillus subtilis] gi|7433750|pir||G69842 (37%
5485 identity in 393 amino acids); its C-terminal-half part is similar to gi|7433750|pir||G69842 (22% identity in 439 amino acids); similar to N- and C -terminal-half part nodulation proteins (nodE) for example ,[Rhizobium meliloti plasmid] gi|128459|sp|P06230|NODE#RHIME, product comprises two
5490 3-oxoacyl-[acyl-carrier- protein]
SEQ ID NO: 1628: -0.368862, 168, novel, similar to(at low level) a part of polyketide synthases for example ,[Streptomyces sp. strain MA6548] gi|7481905|pir||T17428 (23% identity in 201 amino acids)
5495 SEQ ID NO: - : -0.500273, 367, novel
SEQ ID NO: - : -0.253226, 63, a putative ABC transporter ,

- similar to putative ABC transporters (ATP-binding protein)
for example ,[Thermotoga maritima] gi|7445988|pir||H72342
(50% identity in 222 amino acids)
- 5500 SEQ ID NO: 1538: -0.112712, 237, novel
SEQ ID NO: 1539: 0.259358, 188, novel [hypothetical
membrane protein], similar to hypothetical proteins for
example ,BBJ27 [Lyme disease spirochete plasmid J/lp38]
gi|7463605|pir||D70248 (25% identity in 399 amino acids)
- 5505 SEQ ID NO: 1633: -1.014893, 95, novel [periplasmic]
SEQ ID NO: 1634: -0.166975, 325, novel
SEQ ID NO: - : -0.77625, 81, a phage integration, similar to
integrases for example ,[Vibrio cholerae]
gi|498253|gb|aaC44230.1| (32% identity in 390 amino acids)
- 5510 (P4 like integrase)
SEQ ID NO: 2: -0.123944, 214, novel, similar to(at low level)
hemagglutinin main component [Clostridium botulinum phage
(type C)] gi|1346254|sp|P46084|HA33#CLOBO (23% identity
in 190 amino acids)
- 5515 SEQ ID NO: 3: -0.274163, 210, a transposase, similar to sB
proteins for example ,[Shigelladysenteriae Iso-IS1]
gi|6759959|gb|aa F28117.1|AF153317#13 (72% identity in 129
amino acids), GTG start
SEQ ID NO: 4: -0.112565, 192, a putative regulatory protein,
5520 similar to prophage cp4-57regulatory proteinAlpA [Escherichia
coli (strain K-12)] gi|461502|sp|P33997|ALPA#ECOLI (52%
identity in 61 amino acids)
SEQ ID NO: 5: -0.320225, 90, novel, similar to hypothetical
protein b2625 (YfjI) [Escherichia coli K-12]
- 5525 gi|1723621|sp|P52124|YFJI#ECOLI (40% identity in 444
amino acids)
SEQ ID NO: 6: -0.628261, 93, novel, similar to(at low level)
hypothetical protein Cj1244 [Campylobacter jejuni]
gi|6968677|emb|CAB73498.1| (25% identity in 78 amino acids)
- 5530 SEQ ID NO: 7: -0.642435, 272, novel, similar to hypothetical

Appendix B: Hideo *et al.* Full Translation

protein A153R [Chlorella virus PBCV-1]
gi|7461298|pir|T17644 (32% identity in 365 amino acids);
DNA repair protein rad25 PAB0128 [Pyrococcus abyssi (strain
Orsay)]gi|7514780|pir|A75209 (28% identity in 392 amino
5535 acids);putative helicase(D10 protein) [Bacteriophage T5]
gi|137606|sp|P11107|VD10#BPT5 (27% identity in 393 amino
acids)
SEQ ID NO: 8: -0.313568, 200, novel, TTG start
SEQ ID NO: 9: -0.309146, 1160, novel, identical to L0015
5540 [Escherichia coli] gi|3414883|gb|aaC31494.1| (100% identity
in 512 amino acids); similar to hypothetical proteins for
example ,[Escherichia coli] gi|3288156|emb|aa11509.1| (99%
identity in 411 amino acids)
SEQ ID NO: 10: 0.086667, 226, novel, identical to L0014
5545 [Escherichia coli] gi|3288157|emb|Caa11510.1| (100% identity
in 115 amino acids); similar to hypothetical proteins for
example ,orf50 [Escherichia coli] gi|6009426|dbj|Baa84885.1|
(76% identity in 107 amino acids)
SEQ ID NO: 11: -0.430396, 228, novel, similar to hypothetical
5550 proteins for example ,L0013 [Escherichia coli]
gi|3414881|gb|aaC31492.1| (98% identity in 133 amino acids),
GTG start
SEQ ID NO: 12: -0.358621, 233, a IS30 transposase
(interrupted), similar to N-terminal part of IS30 transposas
5555 for example ,i|2851554|sp|P37246|TRA8#ECOLI (99% identity
in 101 amino acids)
SEQ ID NO: 13: -0.43945, 110, a putative transposase,
similar to transposases for example ,Hpl1[Escherichia coli]
gi|3661482|gb|aaC61713.1| (98% identity in 272 amino acids),
5560 InsB [Shigella dysenteriae]
gi|5532467|gb|aaD44751.1|AF141323#22(98% identity in 272
amino acids)
SEQ ID NO: 14: -0.352643, 871, a putative complement
resistance protein precursor, similar to lipoproteinraT

5565 precursors for example ,gi|418135|sp|P32885|TRT1#ECOLI
(83% identity in 227 amino acids), TTG start
SEQ ID NO: 15: -0.186861, 138, novel
SEQ ID NO: 16: -0.535714, 141, novel
SEQ ID NO: 17: -0.34, 251, novel

5570 SEQ ID NO: 18: -0.155725, 132, a putative diacylglycerol
kinase, similar to diacylglycerol kinases for
example ,gi|125321|sp|P00556|KDGL#ECOLI (76% identity in
119 amino acids)
SEQ ID NO: 19: -0.514689, 178, novel[putative outer

5575 membrane protein: OMP], similar to hypothetical proteins
for example ,yjdB in basS-adiY intergenic region
[Escherichiacoli] gi|731986|sp|P30845|YJDB#ECOLI (45%
identity in 428 amino acids)
SEQ ID NO: 20: -0.476923, 118, novel, TTG start

5580 SEQ ID NO: 21: -0.231818, 133, novel
SEQ ID NO: 22: -0.38651, 342, novel, GTG start
SEQ ID NO: 23: -0.159091, 111, an urease accessory protein
UreD, similar to UreD urease-associated proteins for
example ,[Klebsiella aerogenes]

5585 gi|731078|sp|Q09063|URED#KLEAE (71% identity in 242
amino acids), TTG start
SEQ ID NO: 24: -0.048747, 440, an urease gamma subunit,
similar to urease gamma subunits for example ,[Klebsiella
pneumoniae] gi|137084|sp|P18316|URE3#KLEAE (96% identity
in 100 amino acids)

5590 SEQ ID NO: 25: -0.431519, 534, an urease beta subunit,
similar to urease beta subunits for example ,[Klebsiella
pneumoniae] gi|137077|sp|P18315|URE2#KLEAE (82%
identity in 106 amino acids)

5595 SEQ ID NO: 26: -0.526471, 69, an urease alpha subunit,
similar to urease alpha subunits for example ,[Klebsiella
pneumoniae] gi|137070|sp|P18314|URE1#KLEAE (90%
identity in 567 amino acids)

SEQ ID NO: 27: -0.582995, 642, an urease accessory protein,
5600 similar to UreE ureaseaccessory proteins for
example ,[Klebsiella aerogenes]
gi|137095|sp|P18317|UREE#KLEAE (80% identity in 154
amino acids)

SEQ ID NO: 28: -0.439779, 182, an urease accessory protein,
5605 similar to UreF ureaseaccessory protein UreFs for
example ,[Klebsiella aerogenes]
gi|137097|sp|P18318|UREF#KLEAE (79% identity in 224
amino acids)

SEQ ID NO: 29: -0.995946, 75, an urease accessory protein,
5610 similar to UREG urease accessory proteins for
example ,[Klebsiella aerogenes] gi|137099|sp|P18319|UR
EG#KLEAE (90% identity in 205 amino acids)

SEQ ID NO: 30 -0.961539, 105, novel, similar to hypothetical
proteins for example ,TnpJ [Shigella flexneri]
5615 gi|5532468|gb|aaD44752.1|AF141323#23 (100% identity in 87
amino acids)
[0021]

4) Proteins derived from phage

Sequence number: hydrophobicity, The number of amino acids,
5620 Character such as function

SEQ ID NO: 31: 0.178689, 62, a putative antirepressor,
similar to antirepressors for example ,[Bacteriophage 933W]
gi|4585423|gb|aaD25451.1|AF125520#46 (99% identity in 189
amino acids)

5625 SEQ ID NO: 32: -0.403947, 153, a putative host cell lysis,
similar to endolysins for example ,[Bacteriophage 933W]
gi|4585422|gb|aaD25450.1|AF125520#45 (97% identity in 177
amino acids)

SEQ ID NO: 33: -0.280953, 190, novel, similar to hypothetical
5630 protein gi|3183262|sp|P76160|YDFR#ECOLI (45% identity in
74 amino acids)

SEQ ID NO: 34: -0.440678, 178, a putative holin protein,

Appendix B: Hideo *et al.* Full Translation

similar to holins for example ,[Bacteriophage VT2-Sa]
gi|5881636|dbj|Baa84327.1| (97% identity in 68 amino acids)

5635 SEQ ID NO: 35: -0.074561, 115, novel, similar to hypothetical
protein [Bacteriophage VT2-Sa] gi|5881634|dbj|Baa84325.1|
(53% identity in 602 amino acids)
SEQ ID NO: 36: 0.142647, 69, novel, similar to tellurium
resistance protein TerB proteins for example ,[Deinococcus
5640 radiodurans] gi|7473690|pir||C75302 (26% identity in 129
amino acids)
SEQ ID NO: 37: -0.225415, 603, a putative transcription
regulatory element, similar to transcription regulatory
elements for example ,[Escherichia coli]
5645 gi|586679|sp|P37638|YHIW#ECOLI (34% identity in 197
amino acids)
SEQ ID NO: 38: -0.247553, 144, similar to hypothetical
protein [Bacteriophage P27] gi|8346569|emb|CAB93762.1|
(96% identity in 63 amino acids)

5650 SEQ ID NO: 39: 0.054872, 196, a putative anti-terminator
protein, similar to Q protein [Bacteriophage 21]
gi|7440086|pir||D71566 (31% identity in 45 amino acids)
SEQ ID NO: 40: -0.147692, 66, a putative crossover junction
endodeoxyribonuclease, similar to crossover junction
5655 endodeoxyribonuclease [Escherichia coli]
gi|2507117|sp|P40116|RUS#ECOLI (42% identity in 94 amino
acids); Gp67 [Bacteriophage HK97] gi|6901639|gb|aaF31142.1|
(61% identity in 98 amino acids)
SEQ ID NO: 41: -0.278804, 185, similar to B1560#ECOLI
5660 gi|1787843 (85% identity in 354 amino acids)
SEQ ID NO: 42: -0.439604, 102, novel
SEQ ID NO: 43: -0.380555, 361, novel, similar to hypothetical
proteins for example ,[Bacteriophage 933W]
gi|4585451|gb|aaD25479.1|AF125520#74 (99% identity in 114
5665 amino acids); Ygi [Escherichia coli]
gi|1723887|sp|P52083|YGIW#ECOLI (53% identity in 93

amino acids)

SEQ ID NO: 44: -0.741111, 91, a prophage maintenance protein; modulation of host cell killing, identical to MokW

5670 [Bacteriophage 933W]
gi|4585453|gb|aaD25481.1|AF125520#76 (100% identity in 70 amino acids); similar to Hok/Gef family for example ,Gef [Escherichia coli] gi|2120017|pir|S40540 (73% identity in 69 amino acids)

5675 SEQ ID NO: 45: -0.235088, 115, novel, similar to hypothetical proteins for example ,[Bacteriophage 933W]
gi|4585382|gb|aaD25410.1|AF125520#5 (67% identity in 77 amino acids)

5680 SEQ ID NO: 46: 0.222857, 71, novel, similar to hypothetical protein [Bacteriophage 933W]
gi|4585384|gb|aaD25412.1|AF125520#7 (70% identity in 72 amino acids)

SEQ ID NO: 47: -0.37027, 186, novel, GTG start

SEQ ID NO: 48: 0.130555, 73, novel, GTG start

5685 SEQ ID NO: 49: -0.680583, 104, novel, similar to Gp9 [Bacteriophage Mu] gi|6010430|gb|aaF01133.1|AF083977#54 (28% identity in 94 amino acids)

SEQ ID NO: 50: 0.116, 76, novel, similar to hypothetical protein YdaW [Escherichia coli]

5690 gi|3025105|sp|P76066|YDAW#ECOLI, (56% identity in 143 amino acids), TTG start

SEQ ID NO: 51: -0.382796, 94, a putative replication protein, similar to C-terminal-half part of replication protein 14 [Bacteriophage phi-80] gi|137937|sp|P14814|VG14#BPPH8

5695 (45% identity in 129 amino acids)

SEQ ID NO: 52: -0.438934, 245, novel, similar to C-terminal-half part of DnaT [Escherichia coli]
gi|1361001|pir|S56589 (49% identity in 95 amino acids)

SEQ ID NO: 53: -0.760454, 221, novel, similar to hypothetical protein [Escherichia coli] gi|3025103|sp|P76064|YDAT#ECOLI

5700

(30% identity in 141 amino acids)
 SEQ ID NO: 54: -0.684726, 348, a putative regulatory protein,
 similar to Cro [BacteriophageP22]
 gi|132195|sp|P09964|RCRO#BPP22 (39% identity in 53 amino
 5705 acids)
 SEQ ID NO: 55: -0.385816, 142, a putative repressor protein,
 similar to repressor proteins for example ,C2 [Bacteriophage
 P22] gi|133359|sp|P03035|RPC2#BPP22(27% identity in 166
 amino acids)
 5710 SEQ ID NO: 56: -0.0975, 81, novel, similar to hypothetical
 proteins for example ,YdfK[Escherichia coli]
 gi|140584|sp|P29008|YDFA#ECOLI (87% identity in 49 amino
 acids); YdaF gi|3915965|sp|P38395|YDAF#ECOLIF (83%
 identity in 49 amino acids)
 5715 SEQ ID NO: 57: 0.15977, 175, novel, similar to (at low level)
 ATP-dependent protease La homolog
 gi|1708857|sp|P42425|LON2#BACSU (27% identity in 95
 amino acids)
 SEQ ID NO: 58: -0.425974, 78, novel
 5720 SEQ ID NO: 59: -0.477358, 213, novel, TTG start
 SEQ ID NO: 60: -0.526087, 70, a putative cell division
 inhibitor, similar to DicB [Escherichia coli]
 gi|226094|prf||1410309A (67% identity in 55 amino acids)
 SEQ ID NO: 61: -0.439535, 87, novel, similar to hypothetical
 5725 protein YdfD [Escherichia coli]
 gi|140587|sp|P29010|YDFD#ECOLI (45% identity in 62 amino
 acids)
 SEQ ID NO: 62: -0.11129, 63, a putative exonuclease, similar
 to exonucleases for example ,exodeoxyribonuclease VIII
 5730 [Escherichia coli] gi|2507105|sp|P15032|RECE#ECOLI(57%
 identity in 350 amino acids)
 SEQ ID NO: 63: 0.082258, 63, a putative integrase, similar to
 N-terminal part of putative integrases for
 example ,[Escherichia coli cryptic prophage]

5735 gi|7449509|pir||E64913 (93% identity in 183 amino acids),
 TTG start, probably disrupted
 SEQ ID NO: 64: -0.580917, 415, novel
 SEQ ID NO: 65: -0.50929, 184, a transposase (OrfB), identical
 to transposase [Escherichia coli plasmid p O-157 IS629]

5740 gi|7443862|pir||T00240
 SEQ ID NO: 66: -0.175, 85, a transposase (OrfA), identical to
 hypothetical protein [Escherichia coli plasmid p O-157
 intron sequence IS629] gi|7444868|pir||T00241
 SEQ ID NO: 67: -0.397973, 297, a putative transposase,
 similar to putative transposases for example ,[Yersinia
 pestis plasmid pMT1] gi|7447905|pir||T14710 (78%
 identity in 257 amino acids), TTG start, probably disrupted
 SEQ ID NO: 68: -0.965741, 109, novel, identical to L0013
 [Escherichia coli O-157:H7 strain EDL933]

5750 gi|3414881|gb|aaC31492.1| (100% identity in 126 amino
 acids); similar to hypothetical proteins for example ,Hp3
 [Escherichia coli strain CFT073]gi|3661484|gb|aaC61715.1|
 (100% identity in 74 amino acids)
 SEQ ID NO: 69: -0.092042, 290, novel, identical to L0014

5755 [Escherichia coli O-157:H7 strain EDL933]
 gi|3288157|emb|Caa11510.1| (100% identity in 115 amino
 acids); similar to hypothetical proteins for example ,Orf50
 [Escherichia coli strain B171] gi|6009426|dbj|Baa 84885.1|
 (76% identity in 107 amino acids)

5760 SEQ ID NO: 70: -0.403175, 127, novel, identical to L0015
 [Escherichia coli O-157:H7 strain EDL933]
 gi|3414883|gb|aaC31494.1| (100% identity in 512 amino
 acids); similar to hypothetical proteins for
 example ,[Escherichia coli plasmid pColV-K30]

5765 gi|3288156|emb|Caa11509.1| (99% identity in 411 amino
 acids)
 SEQ ID NO: 71: 0.010435, 116, a putative transposase
 (interrupted), similar to N-terminal part of transposases, for

Appendix B: Hideo *et al.* Full Translation

example ,[Escherichia coli strain B171]
5770 gi|1004096|gb|aaB36833.1| (89% identity in 132 amino acids)
SEQ ID NO: 72: -0.445312, 513, novel, similar to hypothetical
proteins for example ,ORF2 in trcA region [Escherichia coli
strain B171-8] gi|4126790|dbj|Baa36748.1| (41% identity in
209 amino acids); ORF4 in trcA region [Escherichia coli strain
5775 B171-8] gi|4126792|dbj|Baa36750.1| (36% identity in 133
amino acids)
SEQ ID NO: 73: -0.736428, 141, novel, similar to hypothetical
protein [Lactococcus bacteriophage c2]
gi|1146281|gb|aaA92162.1| (31% identity in 59 amino acids),
5780 GTG start
SEQ ID NO: 74: -0.321951, 124, novel
SEQ ID NO: 75: -0.187826, 116, novel, similar to hypothetical
proteins for example ,ORF4 in trcA region [Escherichia coli
strain B171-8] gi|4126792|dbj|Baa36750.1| (39% identity in
5785 124 amino acids); ORF2 in trcA region [Escherichia coli strain
B171-8] gi|4126790|dbj|Baa36748.1| (27% identity in 171
amino acids)
SEQ ID NO: 76: 0.102083, 49, novel, similar to hypothetical
protein [Bacteriophage 933W] gi|7649887|dbj|Baa94165.1|
5790 (93% identity in 89 amino acids)
SEQ ID NO: 77: -0.173373, 170, a putative tail fiber protein,
similar to tail fiber proteins for example ,[Bacteriophage
933W] gi|4585436|gb|aaD25464.1|AF125520#59(34% identity
in 339 amino acids)
5795 SEQ ID NO: 78: -0.320225, 90, a putative outer membrane
protein, similar to Lom outer membrane proteins for
example ,[Bacteriophage P-EibA]
gi|7532789|gb|aaF63231.1|AF151091#2 (68% identity in 199
amino acids)
5800 SEQ ID NO: 79: -0.644471, 408, a probably host specificity
protein (partial), similar to C-terminal-half part of protein
J [Bacteriophage lambda]

gi|138412|sp|P03749|VHSJ#LAMB(38% identity in 77 amino acids), GTG start, probably disrupted by frameshift

5805 SEQ ID NO: 80: -0.313568, 200, a host specificity protein (partial), partially similar to protein J [Bacteriophage lambda]

gi|138412|sp|P03749|VHSJ#LAMB(65% identity in 639 amino acids), probably disrupted by frameshift

SEQ ID NO: 81: 0.256338, 72, a host specificity protein

5810 (interrupted), similar to N-terminal part of protein J [Bacteriophage lambda]

gi|138412|sp|P03749|VHSJ#LAMB(80% identity in 369 amino acids), truncated by frameshift

SEQ ID NO: 82: -0.181623, 654, similar to tail assembly, tail assembly proteins for example ,GpI [Bacteriophage lambda]

5815 gi|139637|sp|P03730|VTAI#LAMB(68% identity in 224 amino acids)

SEQ ID NO: 83: -0.403069, 392, tail assembly, similar to tail assembly proteins for example ,GpK [Bacteriophage lambda]

5820 gi|139638|sp|P03729|VTAK#LAMB(85% identity in 196 amino acids), GTG start

SEQ ID NO: 84: 0.103097, 227, a minor tail component, similar to minor tail proteins for example ,GpL [Bacteriophage lambda]

5825 gi|138844|sp|P03738|VMTL#LAMB(76% identity in 232 amino acids)

SEQ ID NO: 85: -0.412946, 225, a putative minor tail component, similar to minor tail proteins for example ,GpM [Bacteriophage lambda]

gi|138845|sp|P03737|VMTM#LAMB(44% identity in 110 amino acids), GTG start

5830 SEQ ID NO: 86: -0.340086, 233, a putative tail length tape measure protein, similar to tail length tape measure proteins for example ,[Bacteriophage HK97]

gi|6901589|gb|aaF31092.1|AF069529#5 (52% identity in 1076 amino acids)

5835 SEQ ID NO: 87: -0.624779, 114, novel, similar to C-terminal

part of Gp14 [Bacteriophage HK97]
gi|6901601|gb|aaF31104.1| (60% identity in 96 amino acids),
probably produced by translational frameshift

5840 SEQ ID NO: 88: -0.311204, 1081, a putative tail assembly
chaperone, similar to tail assembly chaperone [Bacteriophage
HK97] gi|6901600|gb|aaF31103.1| (62% identity in 124 amino
acids)

SEQ ID NO: 89: -0.146237, 94, a putative major tail
5845 component, similar to major tail subunit [Bacteriophage HK97]
gi|6901588|gb|aaF31091.1|AF069529#4 (68% identity in 234
amino acids)

SEQ ID NO: 90: -0.309678, 125, novel, similar to Gp11
[Bacteriophage HK97] gi|6901599|gb|aaF31102.1| (49%
5850 identity in 113 amino acids)

SEQ ID NO: 91: -0.186135, 239, novel, similar to phage
hypothetical protein Gp10 [Bacteriophage HK97]
gi|6901598|gb|aaF31101.1| (75% identity in 148 amino acids)

SEQ ID NO: 92: 0.172807, 115, a putative head-tail adaptor,
5855 similar to putative head-tail adaptors for
example, [Bacteriophage HK97] gi|6901597|gb|aaF31100.1|
(45% identity in 111 amino acids)

SEQ ID NO: 93: -0.512838, 149, novel

SEQ ID NO: 94: -0.192241, 117, a putative portal protein,
5860 similar to portal proteins for example, [Bacteriophage D3]
gi|5059250|gb|aad38955.1| (24% identity in 366 amino acids)

SEQ ID NO: 95: -0.061111, 109, novel

SEQ ID NO: 96: -0.483469, 860, a putative major head
protein/prohead protease, its N-terminal part similar to
5865 putative prohead protease for example, [Rhodobacter
capsulatus] gi|6467535|gb|aaF13181.1|AF181080#3 (30%
identity in 137 amino acids); its C-terminal part similar to
major head proteins for example, [Mycobacterium phage L5]
gi|465114|sp|Q05223|VG17#BPML5 (23% identity in 280
5870 amino acids)

SEQ ID NO: 97: -0.831147, 62, a putative terminase large subunit, similar to hypothetical proteins for example ,phage D3 terminase-like protein [Haemophilus influenzae] gi|6739656|gb|aaF27357.1|AF198256#11 (22% identity in 472 amino acids)

5875 SEQ ID NO: 98: -0.148992, 646, a putative terminase small subunit, similar to terminasesmall subunit - PBSX phage Bacillus subtilis gi|1722886|sp|P39785|XTMA#BACSU (42% identity in 57 amino acids), GTG start

5880 SEQ ID NO: 99: -0.117179, 554, novel

SEQ ID NO: 100: -0.648128, 188, a putative DNase, similar to(at low level) DNase [Bacteriophage phi-C31] gi|1107475|emb|Caa62587.1| (28% identity in 85 amino acids)

SEQ ID NO: 101: -1.029787, 48, novel, similar to hypothetical proteins for example ,[Escherichia coli] gi|1778472|gb|aaB40755.1| (70% identity in 67 amino acids)

5885 SEQ ID NO: 102: -0.468595, 122, a lipoproteinRz1 precursor, similar to lipoproteinRz1 precursors for example ,[Bacteriophage933W]

5890 gi|4585425|gb|aaD25453.1|AF125520#48 (98% identity in 61 amino acids)

SEQ ID NO: 103: -0.717334, 76, an endopeptidase (cell lysis), identical to Rz [Bacteriophage VT2-Sa] gi|5881639|dbj|Baa84330.1|; similar to Rz endopeptidases for

5895 example ,[Bacteriophage lambda] gi|119368|sp|P00726|ENPP#LAMB (69% identity in 153 amino acids)

SEQ ID NO: 104: 0.214754, 62, a putative anti-repressor, identical to Ant [Bacteriophage 933W]

5900 gi|4585423|gb|aaD25451.1|AF125520#46; its N-terminal part (amino acids at the position 1-126) similar to anti-repressor Ant [Bacteriophage P22] gi|131843|sp|P03037|RANT#BPP22 (49% identity in 126 amino acids)

SEQ ID NO: 105: -0.472903, 156, a putative endolysin,

5905 similar to endolysins for example ,[Bacteriophage 933W]
gi|4585422|gb|aaD25450.1|AF125520#45 (96% identity in 177
amino acids)
SEQ ID NO: 106 : -0.283069, 190, novel, similar to
hypothetical protein YdfR (103 amino acids) [Escherichia coli|
5910 gi|3183262|sp|P76160|YDFR#ECOLI (45% identity in 74
amino acids)
SEQ ID NO: 107: -0.466667, 178, a putative holin protein,
similar to holin proteins for example ,[Bacteriophage H-19B]
gi|2668771|gb|aaD04658.1| (97% identity in 68 amino acids)
5915 SEQ ID NO: 108 : -0.074561, 115, novel, similar to
hypothetical proteins for example ,[Bacteriophage 933W]
(52% identity in 613 amino acids)
SEQ ID NO: 109: 0.142647, 69, novel
SEQ ID NO: 110: -0.212987, 617, novel
5920 SEQ ID NO: 111: 0.459524, 43, novel, similar to tellurium
resistance proteins (TerB) for example ,[Deinococcus
radiodurans] gi|7473690|pir||C75302 (26% identity in 120
amino acids), TTG start
SEQ ID NO: 112: -0.452273, 89, novel, TTG start
5925 SEQ ID NO: 113: -0.153521, 143, a putative antitermination
protein, similar to antitermination Q proteins for
example ,[Bacteriophage 82] gi|132277|sp|P13870|RegQ#BP82
(75% identity in 229 amino acids)
SEQ ID NO: 114: -0.142593, 55, a putative crossover junction
5930 endodeoxyribonuclease, similar to Gp67 [Bacteriophage HK97]
gi|6901639|gb|aaF31142.1| (64% identity in 114 amino acids);
crossover junction endodeoxyribonucleases Rus
(Hollidayjunction nuclease) (Holliday junction resolvase)
[Escherichia coli cryptic lambdoid prophage DLP12] (40%
5935 identity in 110 amino acids)
SEQ ID NO: 115: -0.425764, 230, similar to B1560#ECOLI
gi|1787843 (83% identity in 348 amino acids), GTG start
SEQ ID NO: 116: -0.304202, 120, novel

Appendix B: Hideo *et al.* Full Translation

SEQ ID NO: 117: -0.39169, 350, novel

5940 SEQ ID NO: 118: 0.15098, 52, novel

SEQ ID NO: 119: 1.332353, 35, novel, similar to hypothetical protein [Salmonella typhimurium] gi|7467246|pir||T03012 (28% identity in 69 amino acids); Ren proteins for example ,[Bacteriophage H-19] gi|2668762|gb|aaD04649.1|

5945 (26% identity in 109 amino acids)

SEQ ID NO: 120: -0.410309, 195, novel, GTG start

SEQ ID NO: 121: -0.470229, 132, a putative DNA replication protein, similar to DNA replication protein DnaC homologs for example ,[Escherichia coli] gi|7429001|pir||C64886 (79%

5950 identity in 246 amino acids)

SEQ ID NO: 122: -0.365766, 223, a putative replication protein, its C-terminal-half part similar to replication proteins for example ,[Bacteriophage phi-80] gi|137940|sp|P14815|VG15#BPPH8 (34% identity in 148

5955 amino acids); its N-terminal part similar to hypothetical protein [Escherichia coli] gi|3025235|sp|P75978|YMFN#ECOLI (68% identity in 62 amino acids)

SEQ ID NO: 123: -0.47439, 247, novel, similar to hypothetical

5960 protein YdaY [Escherichia coli K-12] gi|3025103|sp|P76064|YDAT#ECOLI (30% identity in 141 amino acids)

SEQ ID NO: 124: -0.667987, 304, novel, similar to hypothetical protein YdaS [Escherichia coli]

5965 gi|3025102|sp|P76063|YDAS#ECOLI (39% identity in 57 amino acids)

SEQ ID NO: 125: -0.42695, 142, novel, similar to hypothetical protein b1145 [Escherichia coli cryptic prophage c14] gi|7444154|pir||F64859 (28% identity in 68 amino acids), TTG

5970 start

SEQ ID NO: 126: -0.183, 101, novel

SEQ ID NO: 127: -0.718055, 145, novel, similar to

hypothetical proteins for example ,[Rhizobium sp. NGR234]
 gi|2496690|sp|P55534|Y4KP#RHISN (38% identity in 89
 5975 amino acids)
 SEQ ID NO: 128: -1.053333, 76, novel
 SEQ ID NO: 129: -0.040217, 93, novel, GTG start
 SEQ ID NO: 130: -0.648148, 55, novel, similar to excisionases
 for example ,[BacteriophageVT2-Sal
 5980 dad|AP000363-2|Baa84285.1| (43% identity in 69 amino acids)
 SEQ ID NO: 131: -0.001695, 119, novel [hypothetical
 lipoprotein], similar to hypothetical proteins for
 example ,CJ0034c [Campylobacter jejuni]
 gi|6967539|emb|CAB72527.1 (35% identity in 229 amino acids),
 5985 GTG start
 SEQ ID NO: 1595: -0.731325, 84, a transposase (insertion
 sequence IS629), similar to hypothetical proteins for
 example ,TnpE [Shigella flexneri]
 gi|5532454|gb|aaD44738.1|AF141323#9 (99% identity in 108
 5990 amino acids)
 SEQ ID NO: 1684: -0.126695, 237, a transposase (OrfB)
 (insertion sequenceIS629), similar to transposase IS629
 gi|7443863|pir|T00315 (98% identity in 295 amino acids)
 SEQ ID NO: 1647: -0.938889, 109, a putative integrase,
 5995 similar to integrases for example ,[Bacteriophage S2]
 gi|1679807|emb|Caa96221.1| (57% identity in 331 amino
 acids)
 SEQ ID NO: 1648: -0.432542, 296, novel, similar to(at low
 level) hypothetical protein b1839[Escherichia coli]
 6000 gi|7451973|pir|G64945 (33% identity in 109 amino acids)
 SEQ ID NO: 1158: -0.498198, 334, novel, similar to(at low
 level) cell division protein Div [Escherichia coli]
 gi|2507010|sp|P15286 (27% identity in 121 amino acids)
 SEQ ID NO: 1159: -0.102609, 116, a putative transcription
 6005 regulatory element, similar to putative transcription
 regulatory elements for example ,[Neisseria meningitidis]

Appendix B: Hideo *et al.* Full Translation

gi|7226247|gb|aaF41408.1| (32% identity in 102 amino acids)
 SEQ ID NO: 1160: -0.209722, 217, novel
 SEQ ID NO: 1161: -0.639552, 135, a putative DNA-binding
 6010 protein, similar to putative DNA-binding protein Cox [Vibrio
 cholerae Bacteriophage K139] gi|4530499|gb|aaD22064.1|
 (46% identity in 56 amino acids); phage hypothetical proteins
 for example ,[Bacteriophage S2] gi|1679810|emb|Caa 96224.1|
 (42% identity in 61 amino acids); [Escherichia coli retron EC67]
 6015 gi|141342|sp|P21315|YR7A#ECOLI (42% identity in 61 amino
 acids)
 SEQ ID NO: 1162: -0.051111, 46, novel
 SEQ ID NO: 1163: 0.01194, 68, novel
 SEQ ID NO: 1164: -0.692241, 117, novel
 6020 SEQ ID NO: 1165: -0.229348, 93, novel
 SEQ ID NO: 1166: -0.27625, 81, novel
 SEQ ID NO: 1167: -0.094928, 139, novel
 SEQ ID NO: 1168: -0.673134, 68, novel
 SEQ ID NO: 1169 : -0.281818, 89, novel, similar to
 6025 hypothetical proteins for example ,[Shigella flexneri]
 gi|421263|pir||S34345 (41% identity in 84 amino acids)
 SEQ ID NO: 1170: -0.030303, 100, a putative derepression
 protein, similar to(at low level) derepression protein epsilon
 [Bacteriophage P4] gi|137833|sp|P05463|VEPS#BPP4 (32%
 6030 identity in 50 amino acids)
 SEQ ID NO: 1171: -0.201464, 206, novel
 SEQ ID NO: 1172 : -0.709211, 77, a putative replication
 protein, similar to replication proteins for example ,GpA
 [Bacteriophage 186] gi|1351406|sp|P41064|VPA#BP186 (34%
 6035 identity in 567 amino acids)
 SEQ ID NO: 1173 : -0.276033, 122, putative regulation of
 plasmid partition, similar to plasmid partition proteins for
 example ,par [Escherichia coli plasmid R1]
 gi|134954|sp|P11904|STBA#ECOLI (46% identity in 314
 6040 amino acids)

SEQ ID NO: 1174: -0.74575, 895, regulation of plasmid partition, similar to plasmid partition proteins for example ,TSB [Escherichia coli plasmid NR1|gi|134956|sp|P11906|STBB#ECOLI (40% identity in 62 amino acids)

6045

SEQ ID NO: 1175: -0.094984, 320, a putative transposase, its N-terminal part (amino acids at the position 1-103/217) is identical to N-terminal part of transposase [Escherichia coli plasmid p O-157 insertion sequence IS629|gi|7443862|pir||T00240(1-103/296 amino acids), its C-terminal part (amino acids at the position 104-217/217) is identical to C-terminal part of transposase [Escherichia coli plasmid p O-157 insertion sequence IS629|gi|7443862|pir||T00240(183-296/296 amino acids)

6050

6055

SEQ ID NO: 1176: -0.466346, 105, a transposase, similar to hypothetical proteins in insertion sequences for example ,[Escherichia coli plasmid p O-157 insertion sequence IS629|gi|7444868|pir||T00241 (96% identity in 108 amino acids)

6060

SEQ ID NO: 1177 : -0.368996, 230, novel, similar to hypothetical proteins for example ,orf20 [Escherichia coli plasmid pB171|gi|6009396|dbj|Baa84855.1| (54% identity in 158 amino acids) (transferase)

6065

SEQ ID NO: 1178: -0.912037, 109, a putative tail protein, similar to tail proteins for example ,F protein [Bacteriophage 186|gi|3337273|gb|aaC34171.1| (43% identity in 151 amino acids)

6070

SEQ ID NO: 1179 : -0.174684, 159, novel, similar to C-terminal part of tail proteins for example ,GpT [Bacteriophage P2|gi|3139112|gb|aaD03293.1|(39% identity in 66 amino acids), GTG start, probably disrupted by frameshift

SEQ ID NO: 1180: -0.337037, 163, a putative tail protein, similar to N-terminal part of tail proteins for example ,GpT [Bacteriophage P2|gi|3337272|gb|aaC34170.1| (32% identity

6075 in 648 amino acids), interrupted by frameshift
 SEQ ID NO: 1181: -0.326978, 279, a putative phage tail
 protein, similar to gi|3139111|gb|aaD03292.1| (47% identity
 in 42 amino acids)
 SEQ ID NO: 1182: -0.055746, 697, a putative tail protein,
 6080 similar to tail proteins for example ,GpE [Bacteriophage P2]
 gi|3139110|gb|aaD03291.1| (31% identity in 85 amino acids)
 SEQ ID NO: 1183: -0.129487, 79, a putative tail tube
 protein, similar to tail tube proteins for example ,tail
 protein FII [Bacteriophage 186]
 6085 gi|139325|sp|P22502|VPF2#BPP2 (44% identity in 157 amino
 acids)
 SEQ ID NO: 1184: -0.284298, 122, a putative tail sheath
 protein, similar to tail sheath proteins for example ,FI
 [Pseudomonas aeruginosa bacteriophage phiCTX]
 6090 gi|4063795|dbj|Baa36249.1| (47% identity in 377 amino acids)
 SEQ ID NO: 1185: -0.266471, 171, a tail protein, similar to
 N-terminal part of tail proteins for example ,GpD
 [Bacteriophage P2] gi|6136287|sp|P10312|VPD#BPP2 (59%
 identity in 70 amino acids)
 6095 SEQ ID NO: 1186: -0.193147, 395, a transposase, similar to
 transposases for example ,[Escherichia coli insertion sequence
 IS30] gi|2851554|sp|P37246|TRA8#ECOLI (99% identity in
 342 amino acids)
 SEQ ID NO: 1187: -0.173832, 108, novel, GTG start
 6100 SEQ ID NO: 1188: -0.841108, 344, novel
 SEQ ID NO: 1189: -0.626563, 65, similar to FLIC#ECOLI
 gi|1788232 (55% identity in 585 amino acids)
 SEQ ID NO: 1190: -0.435484, 94, its N-terminal part (amino
 acids at the position 1-104/379) similar to YEDM#ECOLI
 6105 gi|1788245 (77% identity in 104 amino acids), its central part
 (amino acids at the position 162-266/379) is similar to
 YEDN#ECOLI gi|1788244 (60% identity in 105 amino acids), its
 C-terminal part (amino acids at the position 272-331/379) is

- similar to B1933#ECOLI gi|1788243 (46% identity in 59 amino acids) ; similar to(at low level) YOPM#YERPE SP|P17778 (27% identity in 181 amino acids)
- 6110 SEQ ID NO: - : -0.296752, 586, similar to C-terminal part of YEDL#ECOLI gi|1788242 (61-159/159 amino acids) (93% identity in 99 amino acids)
- 6115 SEQ ID NO: - : -0.242216, 380, its N-terminal part (amino acids at the position 1-104/379) is similar to YEDM#ECOLI gi|1788245 (76% identity in 104 amino acids), its central part (amino acids at the position 162-266/379) is similar to YEDN#ECOLI gi|1788244 (61% identity in 105 amino acids), its
- 6120 C-terminal part (amino acids at the position 272-331/379) is similar to B1933#ECOLI gi|1788243 (53% identity in 59 amino acids); similar to(at low level) IPA#SHIFL dad|M32063-1 (30% identity in 146 amino acids)
- SEQ ID NO: 1554 -0.263636, 100, novel, TTG start
- 6125 SEQ ID NO: - : -0.244327, 380, novel
- SEQ ID NO: - : -0.468966, 117, a putative secreted effector protein, similar to hypothetical proteins for example ,EspF [Escherichia coli strain B10] gi|6090818|gb|aaF03351.1|AF116900#2 ESPF#ECOLI (39% identity in 126 amino acids)
- 6130 SEQ ID NO: 756 : -0.497235, 218, novel, similar to hypothetical protein [Bacteriophage 933W] gi|4585437|gb|aaD25465.1|AF125520#60 (93% identity in 89 amino acids)
- 6135 SEQ ID NO: 757: -0.686944, 338, a putative bacteriophage tail fiber protein, similar to tail fiber proteins for example ,[Bacteriophage 933W] gi|4585436|gb|aaD25464.1|AF125520#59 (38% identity in 370 amino acids)
- 6140 SEQ ID NO: 758: -0.324719, 90, a putative outer membrane protein, similar to Lom outer membrane protein precursors for example ,[prophage P-EibA]

gi|7532789|gb|aaF63231.1|AF151091#2 (68% identity in 199 amino acids)

6145 SEQ ID NO: 759 : -0.67254, 438, a bacteriophage host specificity protein(partial), similar to C-terminal part of host specificity proteins for example ,GpJ [Bacteriophage lambda|gi|138412|sp|P03749|VHSJ#LAMBD (58% identity in 788 amino acids), probably disrupted by frameshift

6150 SEQ ID NO: 760 : -0.313568, 200, a bacteriophage host specificity protein (interrupted), similar to N-terminal part of host specificity proteins for example , protein J [Bacteriophage lambda| gi|138412|sp|P03749|(80% identity in 369 amino acids), GTG start, interrupted by frameshift

6155 SEQ ID NO: 761: -0.245668, 809, a putative tail assembly protein, similar to tail assembly proteins for example ,GpI [Bacteriophage lambda| gi|139637|sp|P03730|VTAI#LAMBD (69% identity in 224 amino acids)

SEQ ID NO: 762: -0.365217, 392, bacteriophage tail assembly, similar to tail assembly proteins for example ,GpK [Bacteriophage lambda| gi|139638|sp|P03729|VTAK#LAMBD (87% identity in 186 amino acids)

6160 SEQ ID NO: 763: 0.086667, 226, a possible bacteriophage tail component, similar to minor tail proteins for example ,GpL [Bacteriophage lambda| gi|138844|sp|P03738|VMTL#LAMBD (76% identity in 232 amino acids)

6165 SEQ ID NO: 764 : -0.344973, 190, a bacteriophage tail component, similar to minor tail proteins for example ,GpM [Bacteriophage lambda| gi|138845|sp|P03737|VMTM#LAMBD (79% identity in 109 amino acids)

6170 SEQ ID NO: 765 : -0.3125, 233, tail length determination, similar to C-terminal part of tail length tape measure protein precursors for example ,GpH [Bacteriophage lambda| gi|138843|sp|P03736|VMTH#LAMBD (80% identity in 253 amino acids), probablydisrupted by frameshift

6175 SEQ ID NO: 766 : -0.43945, 110, bacteriophage tail length

determination, similar to N-terminal part tail length tape
measure proteins for example ,GpH [Bacteriophage lambda]
gi|138843|sp|P03736|VMTH#LAMBDA (76% identity in 587
6180 amino acids), interrupted by frameshift
SEQ ID NO: 767 : -0.258268, 255, a bacteriophage tail
component, similar to minor tail proteins for example ,GpT
[Bacteriophage lambda] gi|138846|sp|P03735|VMTT#LAMBDA
(78% identity in 96 amino acids), probably produced by
6185 translationframeshift
SEQ ID NO: 768: -0.505, 621, a bacteriophage tail component,
similar to minor tail proteins for example ,GpG
[Bacteriophage
lambda]gi|138842|sp|P03734|VMTG#LAMBDA(68% identity in
6190 167 amino acids)
SEQ ID NO: 769: 0.034653, 102, novel
SEQ ID NO: 770 : -0.22028, 144, a bacteriophage head
component, similar to N-terminal part of major head proteins
for example ,Gp7 [Bacteriophage 21]
6195 gi|547612|sp|P36270|HEAD#BPP21percent
gi|547612|sp|P36270| (95% identity in 88 amino acids),
probably interrupted
SEQ ID NO: 771 : -0.239801, 202, a bacteriophage head
component, similar to head decoration proteins for
6200 example ,Gpshp [Bacteriophage 21]
gi|549437|sp|P36275|VSHP#BPP21 (95% identity in 115
amino acids)
SEQ ID NO: 772 : -0.331818, 89, a bacteriophage head-tail
preconnector, similar to minor head proteins for
6205 example ,head-tail preconnector Gp5 [Bacteriophage 21]
gi|549296|sp|P36273|VG05#BPP21 (97% identity in 501 amino
acids), scaffold protein(302-501 amino acids) containing
homolog of Gp6 [Bacteriophage 21]
SEQ ID NO: 773 : -0.024348, 116, a bactreiophage portal
6210 protein, similar to portal proteins for example ,Gp5

[Bacteriophage 21] gi|549295|sp|P36272|VG04#BPP21 (98% identity in 530 amino acids)

SEQ ID NO: 774: 0.055688, 502, a putative head completion protein, similar to phage proteins for example ,head completion protein Gp3 [Bacteriophage 21] 6215 gi|549294|sp|P36271|VG03#BPP21 (98% identity in 68 amino acids)

SEQ ID NO: 775: -0.448868, 531, a bacteriophage terminase large subunit, similar to terminase large subunits for example ,Gp2 [Bacteriophage 21] 6220 gi|2851579|sp|P36693|TERL#BPP21 (91% identity in 637 amino acids)

SEQ ID NO: 776: -0.394118, 69, a possible bacteriophage terminase small subunit, similar to terminase small subunits for example ,Gp1 [Bacteriophage N15] gi|7444578|pir|T13087 6225 (42% identity in 106 amino acids), GTG start

SEQ ID NO: 777: -0.425233, 643, a transcription regulatory element, similar to PerC (BfpW) [Escherichiacoli] gi|1172431|sp|P43475|PERC#ECOLI (47% identity in 87 6230 amino acids)

SEQ ID NO: 778: -0.508875, 170, a lipoprotein precursor, similar to lipoproteinRz1 precursors for example ,[Bacteriophage 933W] gi|4585425|gb|aaD25453.1|AF125520#48 (85% identity in 61 6235 amino acids)

SEQ ID NO: 779: -0.983654, 105, an endopeptidase (host cell lysis), similar to Rzendopeptidases for example ,[Bacteriophage VT2-Sal] gi|5881639|dbj|Baa84330.1|(83% identity in 154 amino acids)

6240 SEQ ID NO: 780: 0.178689, 62, novel

SEQ ID NO: 781: -0.26, 156, similar to possible endolysins, for example ,R protein [Bacteriophage H-19B] gi|4335686|gb|aaD17382.1| (98% identity in 177 amino acids)

SEQ ID NO: 782: 0.62, 61, novel, similar to Ydfr [Escherichia

6245 coli| gi|3183262|sp|P76160|YDFR#ECOLI (44% identity in 74 amino acids)
 SEQ ID NO: 783: -0.393785, 178, a possible holin protein (host cell lysis), similar to holin proteins for example , protein [Bacteriophage VT2-Sa| gi|5881636|dbj|Baa84327.1| (94% identity in 68 amino acids)

6250 SEQ ID NO: 784: -0.114912, 115, a transposase, identical to hypothetical protein[Escherichia coli plasmid p O-157 insertion sequence IS629| gi|7444868|pir||T00241
 SEQ ID NO: 785: 0.133823, 69, a transposase, identical to transposase [Escherichia coli plasmid p O-157 insertion sequence IS629| gi|7443862|pir||T00240

6255 SEQ ID NO: 786 : -0.965741, 109, novel, similar to hypothetical proteins for example ,[Bacteriophage 933W] gi|4585419|gb|aaD25447.1|AF125520#42 (53% identity in 613 amino acids)

6260 SEQ ID NO: 787: -0.397973, 297, novel, GTG start
 SEQ ID NO: 788: -0.243181, 617, novel, GTG start
 SEQ ID NO: 789: 0.475926, 55, novel, similar to putative TerB proteins for example ,[Deinococcus radiodurans] gi|7473690|pir||C75302 (26% identity in 120 amino acids)

6265 SEQ ID NO: 790: 1.385455, 56, an antitermination, similar to antiterminators for example , protein Q [Bacteriophage 82] gi|132277|sp|P13870|R for example ,Q#BP82 (75% identity in 229 amino acids)

6270 SEQ ID NO: 791 : -0.143662, 143, a crossover junction endodeoxyribonuclease, similar to Rus proteins for example ,[Bacteriophage 82| gi|6901639|gb|aaF31142.1| GP67#BPHK97 (63% identity in 112 amino acids); similar to Gp67 [Bacteriophage HK97] gi|6901639|gb|aaF31142.1| (63% identity in 112 amino acids)

6275 SEQ ID NO: 792 : -0.393886, 230, novel, similar to hypothetical proteins for example ,b1560 [Escherichia coli] gi|7466196|pir||C64911 (85% identity in 348 amino acids),

GTG start

- 6280 SEQ ID NO: 793: -0.221009, 120, novel, similar to orf QD1
[Bacteriophage N15] gi|2564084|gb|aaB81659.1| (31% identity
in 64 amino acids)
SEQ ID NO: 794: -0.35702, 350, a prophage maintenance
(modulation of hostcell killing), similar to Hok/Gef family for
6285 example ,MokW [Bacteriophage 933W]
gi|4585453|gb|aaD25481.1|AF125520#76 (87% identity 70
amino acids)
SEQ ID NO: 795: -1.208696, 93, novel
SEQ ID NO: 796: 0.081429, 71, novel, its N-terminal part
6290 (amino acids at the position 1-46 amino acids) is similar to
GP45 [Bacteriophage N15] gi|7521552|pir|T13131 (56%
identity in 46 amino acids); its N-terminal part (amino acids at
the position 37-97) is similar to b2363 [Escherichia coli]
gi|7451977|pir|H65009 (73% identity in 61 amino acids)
6295 SEQ ID NO: 797: 1.402941, 35, novel
SEQ ID NO: 798: -0.425134, 188, novel, GTG start
SEQ ID NO: 799: -0.893204, 104, novel
SEQ ID NO: 800: -1.069355, 63, novel
SEQ ID NO: 801: -0.171186, 119, novel, similar to YdaW
6300 [Escherichia coli] gi|3025105|sp|P76066|YDAW#ECOLI (61%
identity in 135 amino acids)
SEQ ID NO: 802: -0.148649, 75, a putative phage replication
protein, similar to replication proteins for example ,Gp14
[Bacteriophage phi-80] gi|137937|sp|P14814|VG14#BPPH8
6305 (47% identity in 129 amino acids)
SEQ ID NO: 803: -0.504741, 233, novel, similar to replication
termination factor dnaT (primosomal protein I) [Escherichia
coli] gi|1361001|pir|S56589 (30% identity in 85 amino acids)
SEQ ID NO: 804: -0.721364, 221, novel, similar to YdaT
6310 [Escherichia coli] gi|3025103|sp|P76064|YDAT#ECOLI (31%
identity in 83 amino acids); similar to(at low level) regulatory
protein CII [Bacteriophage phi-80]

Appendix B: Hideo *et al.* Full Translation

- gi|133360|sp|P14820|RPC2#BPPH8 (40% identity in 40 amino acids)
- 6315 SEQ ID NO: 805: -0.660869, 346, a putative cell division control protein (repressor), similar to DicC (repressor protein of division inhibition genedicB) [Escherichia coli|gi|118633|sp|P06965|DICC#ECOLI (31% identity in 72 amino acids)
- 6320 SEQ ID NO: 806: -0.360284, 142, a possible repressor protein, similar to repressor proteins for example ,C2 repressor [Bacteriophage P22|gi|133359|sp|P03035|RPC2#BPP22 (30% identity in 203 amino acids)
- SEQ ID NO: 807: -0.694667, 76, novel
- 6325 SEQ ID NO: 808: -0.046047, 216, a possible cell division inhibitor, similar to DicB protein [Escherichia coli|gi|2507009|sp|P09557|DICB#ECOLI (65% identity in 55 amino acids)
- SEQ ID NO: 809: -0.494, 51, novel, similar to hypothetical proteins for example ,YdfD[Escherichia coli|gi|140587|sp|P29010|YDFD#ECOLI (46% identity in 62 amino acids)
- 6330 SEQ ID NO: 810: -0.01129, 63, novel
- SEQ ID NO: 811: 0.119355, 63, novel
- 6335 SEQ ID NO: 812: -0.751913, 733, novel
- SEQ ID NO: 813: -0.487736, 107, an integrase, similar to integrases for example ,[Bacteriophage HK022]|gi|138560|sp|P16407|VINT#BPHK0 (24% identity in 316 amino acids)
- 6340 SEQ ID NO: 814: -0.347761, 68, novel, similar to hypothetical proteins for example ,L0013 [Escherichia coli O-157:H7 strain EDL933|gi|3414881|gb|aaC31492.1|(100% identity in 133 amino acids), GTG start
- SEQ ID NO: 815: -0.722352, 341, novel, similar to
- 6345 hypothetical proteins for example ,L0014 [Escherichia coli O-157:H7 strain EDL933|gi|3288157|emb|Caa11510.1| (100%

identity in 115 amino acids)
 SEQ ID NO: 1581 : -0.388722, 134, novel, similar to
 hypothetical proteins for example ,L0015 [Escherichia coli
 6350 O-157:H7 strain EDL933] gi|3414883|gb|aaC31494.1| (100%
 identity in 512 amino acids)
 SEQ ID NO: 1582: 0.010435, 116, novel
 SEQ ID NO: 1583: -0.445312, 513, a transposase (insertion
 sequence IS629), similar to IS629 hypothetical proteins for
 6355 example ,[Escherichia coli plasmid p O-157]
 gi|7444868|pir||T00241 (96% identity in 108 amino acids)
 SEQ ID NO: 1349 : -0.262963, 55, a transposase (insertion
 sequence IS629), similar to IS629 transposase [Escherichia coli
 plasmid p O-157] gi|7443862|pir||T00240 (96% identity in
 6360 296 amino acids)
 SEQ ID NO: 1350: -0.942593, 109, novel, partially similar
 to hypothetical proteins for example ,YjdA [Escherichia coli]
 gi|731985|sp|P16694|YJDA#ECOLI (17% identity in 236
 amino acids) (at low level)
 6365 SEQ ID NO: 1351 : -0.402027, 297, novel, similar to
 hypothetical protein YjcZ [Escherichia coli]
 gi|731984|sp|P39267|YJCZ#ECOLI (29% identity in 278 amino
 acids), GTG start
 SEQ ID NO: 1352: -0.652559, 294, novel, similar to (at low
 6370 level) hypothetical proteins for example ,[Xanthomonas
 campestris] gi|6689533|emb|CAB65709.1| (44% identity in 74
 amino acids)
 SEQ ID NO: 1353: -0.372093, 302, novel
 SEQ ID NO: 1354: 0.036798, 357, novel
 6375 SEQ ID NO: 1355 : -0.067841, 228, novel, similar to
 hypothetical proteins for example ,YafZ [Escherichia coli]
 gi|2495487|sp|P77206|YAFZ#ECOLI (75% identity in 272
 amino acids)
 SEQ ID NO: 1356: -0.074265, 137, a putative antirestriction
 6380 protein, similar to hypothetical proteins for example ,YfjX

[*Escherichia coli*] gi|1723636|sp|P52139|YFJX#ECOLI (68% identity in 152 amino acids); similar to antirestriction proteins for example ,KlcA protein | plasmid RK2| gi|1730051|sp|P52603|KLA2#ECOLI (38% identity in 139 amino acids)

6385 SEQ ID NO: 1357 : -0.550183, 274, an acetyltransferase, identical to WbdR [*Escherichia coli* O-157:H7 C664-1992] gi|3435182|gb|aaC32350.1|

6390 SEQ ID NO: 1358 : -0.385535, 160, novel, similar to C-terminal part of H repeat-associated proteins for example ,[*Escherichia coli*] gi|140772|sp|P28912|YHHI#ECOLI (66% identity in 36 amino acids), TTG start

6395 SEQ ID NO: 1259 : 0.180543, 222, novel, similar to H repeat-associated proteins for example ,[*Escherichia coli*] gi|140772|sp|P28912|YHHI#ECOLI (75% identity in 49 amino acids)

6400 SEQ ID NO: 1260 : 0.204, 51, novel, similar to H repeat-associated proteins for example ,[*Escherichia coli*] gi|140772|sp|P28912|YHHI#ECOLI (83% identity in 36 amino acids), GTG start

6405 SEQ ID NO: 1261 : -0.351852, 55, a phosphomannomutase, identical to ManB [*Escherichia coli* O-157:H7 C664-1992] gi|3435181|gb|aaC32349.1|

6410 SEQ ID NO: 1262 : -0.141667, 37, a mannose-1-P guanosyltransferase, identical to ManC [*Escherichia coli* O-157:H7 C664-1992] gi|3435180|gb|aaC32348.1|

SEQ ID NO: 1263 : -0.222368, 457, a probable GDP-L-fucose pathway enzyme, identical to WbdQ [*Escherichia coli* O-157:H7 C664-1992] gi|3435179|gb|aaC32347.1|

SEQ ID NO: 1264 : -0.221577, 483, a fucose synthetase, identical to Fcl [*Escherichia coli* O-157:H7 C664-1992] gi|4867922|dbj|Baa77731.1|

SEQ ID NO: 1265 : -0.168047, 170, a GDP-D-mannose

Appendix B: Hideo *et al.* Full Translation

6415 dehydratase, identical to Gmd [Escherichia coli O-157:H7 C664-1992] gi|3435177|gb|aaC32345.1|
 SEQ ID NO: 1266: -0.264486, 322, a (e) glycosyl transferase, similar to WbdP [Escherichia coli O-157:H7 C664-1992] gi|3435176|gb|aaC32344.1|

6420 SEQ ID NO: 1267: -0.261021, 373, a perosamine synthetase, identical to Per [Escherichia coli O-157:H7 C664-1992] gi|3435175|gb|aaC32343.1|
 SEQ ID NO: 1268: -0.176485, 405, an O antigen flippase, identical to Wzx [Escherichia coli O-157:H7 C664-1992] gi|3435174|gb|aaC32342.1|

6425 SEQ ID NO: 1269: -0.321585, 367, a probable glycosyl transferase, identical to WbdO [Escherichia coli O-157:H7 C664-1992] gi|3435173|gb|aaC32341.1|
 SEQ ID NO: 1270: 0.75141, 462, an O antigen polymerase, identical to Wzy [Escherichia coli O-157:H7 C664-1992] gi|3435172|gb|aaC32340.1|, GTG start
 SEQ ID NO: 1271: -0.16371, 249, a (e) glycosyl transferase, identical to WbdN [Escherichia coli O-157:H7 C664-1992] gi|4867915|dbj|Baa77724.1|

6435 SEQ ID NO: 1272: 0.558884, 395, a putative UDP-galactose 4-epimerase, similar to putative UDP-galactose 4-epimerase [Vibrio cholerae] gi|3724321|dbj|Baa33610.1 (27% identity in 329 amino acids)
 SEQ ID NO: 1273: -0.404615, 261, novel, similar to hypothetical proteins for example ,gi|9106618|gb|aaF84382.1|AE003986#12 [Xylella fastidiosa] (60% identity in 105 amino acids)
 SEQ ID NO: 1638: -0.29577, 332, novel, similar to hypothetical protein [Xylella fastidiosa] gb|aaF84486.1|AE003993#5 (52% identity in 86 amino acids)

6445 SEQ ID NO: 1692: -0.842857, 113, novel
 SEQ ID NO: 1693: -0.109375, 97, novel
 SEQ ID NO: 1588: -0.478481, 80, novel [putative outer

membrane protein; OMP]

6450 SEQ ID NO: 1589: -0.057391, 116, similar to YEHA#ECOLI
gi|1788426 (44% identity in 207 amino acids) [putative type-1
fimbrial protein]

SEQ ID NO: 1590: 0.006731, 105, similar to YEHB#ECOLI
gi|1788427 (92% identity in 826 amino acids); similar to usher

6455 protein MrkC [Klebsiella pneumoniae]
dad|M55912-4|aaA25095.1 (32% identity in 810 amino acids)

SEQ ID NO: - : -0.098256, 345, similar to YEHC#ECOLI
gi|1788428 (87% identity in 224 amino acids); similar to
chaperone MrkB [Klebsiella pneumoniae]

6460 dad|M55912-3|aaA25094.1 (39% identity in 211 amino acids)

SEQ ID NO: - : -0.513075, 827, similar to YEHD#ECOLI
gi|1788429 (85% identity in 180 amino acids); AC/I pili
protein [Escherichia coli] dad|X76121-1|Caa53727.1 (28%
identity in 177 amino acids)

6465 SEQ ID NO: - : -0.266071, 225, similar to YEHE#ECOLI
gi|1788430 (69% identity in 93 amino acids)

SEQ ID NO: - : 0.199444, 181, a putative molybdate
metabolism regulator, similar to N-terminal part of molybdate
metabolism regulator MolR [Escherichia coli]

6470 gi|7466653|pir||B64979(amino acids at the position
1-244/1264) (37% identity in 249 amino acids), GTG start

SEQ ID NO: - : -0.272043, 94, a putative molybdate
metabolism regulator, similar to C-terminal part of molybdate
metabolism regulator molR [Escherichia coli]

6475 gi|465576|sp|P33345|MOLR#ECOLI (45% identity in 1000
amino acids), GTG start

SEQ ID NO: - : -0.647107, 243, identical to transposase (OrfB)
(insertion sequence IS629), gi|7443862|pir||T00240

SEQ ID NO: 1509: -0.306124, 948, similar to transposase
(OrfA) (insertion sequence IS629), gi|7444868|pir||T00241
(99% identity in 108 amino acids)

SEQ ID NO: 1650: -0.397973, 297, novel

Appendix B: Hideo *et al.* Full Translation

SEQ ID NO: 1651 : -0.958333, 109, novel, similar to hypothetical proteins for example ,[Bacteriophage 933W]
6485 gi|4585437|gb|aaD25465.1|AF125520#60 (97% identity in 102 amino acids), TTG start

SEQ ID NO: 555: -0.584146, 83, a putative tail fiber protein, similar to tail fiber proteins for example ,[Bacteriophage 933W]
6490 gi|4585436|gb|aaD25464.1|AF125520#59 (36% identity in 361 amino acids)

SEQ ID NO: 556: -0.411765, 103, a putative outer membrane protein Lom precursor, similar to Lom precursors for example ,[Bacteriophage P-EibA]
6495 gi|7532789|gb|aaF63231.1|AF151091#2 (76% identity in 199 amino acids)

SEQ ID NO: 557: -0.679634, 438, a putative host specificity protein (partial), similar to C-terminal part of host specific proteins for example ,GpJ [Bacteriophage lambda]
6500 gi|138412|sp|P03749|VHSJ#LAMB(62% identity in 775 amino acids), GTG start

SEQ ID NO: 558: -0.288442, 200, a putative host specific protein (interrupted), similar to N-terminus of host specificity proteins for example ,GpJ [Bacteriophage lambda]
6505 gi|138412|sp|P03749|VHSJ#LAMB(80% identity in 369 amino acids), GTG start, probably truncated by framesift

SEQ ID NO: 559: -0.197032, 776, a putative tail assembly protein, similar to tail assembly proteins for example ,GpI [Bacteriophage lambda] gi|139637|sp|P03730|VTAI#LAMB(69% identity in 224 amino acids)

6510 SEQ ID NO: 560: -0.365217, 392, a putative tail assembly protein, similar to tail assembly proteins for example ,GpK [Bacteriophage lambda] gi|139638|sp|P03729|VTAK#LAMB(86% identity in 196 amino acids)

SEQ ID NO: 561: 0.086667, 226, a putative minor tail protein, similar to minor tail proteins for example ,GpII[Bacteriophage lambda]

Appendix B: Hideo *et al.* Full Translation

gi|138844|sp|P03738|VMTL#LAMBD (76% identity in 232 amino acids)
 SEQ ID NO: 562: -0.32996, 248, a putative minor tail protein,
 6520 similar to minor tail proteins for example ,GpM [Bacteriophage lambda]
 gi|138845|sp|P03737|VMTM#LAMBD(79% identity in 109 amino acids)
 SEQ ID NO: 563: -0.3125, 233, a putative tail length tape
 6525 measure protein precursor, similar to tail length tape measure protein precursors for example ,GpH [Bacteriophagelambda]
 gi|138843|sp|P03736|VMTH#LAMBD (49% identity in 876 amino acids)
 SEQ ID NO: 564: -0.43945, 110, a putative minor tail
 6530 protein, similar to minor tail proteins for example ,GpT [Bacteriophage lambda]
 gi|138846|sp|P03735|VMTT#LAMBD(70% identity in 102 amino acids), probably produced by translational frameshift
 SEQ ID NO: 565: -0.353916, 882, a putative minor tail
 6535 protein, similar to minor tail proteins for example ,GpG [Bacteriophage lambda] gi|138842|sp|P03734|VMTG#LAMBD (43% identity in 140 amino acids)
 SEQ ID NO: 566: -0.358824, 103, novel
 SEQ ID NO: 567: -0.545714, 141, a putative minor tail
 6540 protein U, similar to minor tail proteins for example ,GpU [Bacteriophage lambda] gi|138847|sp|P03732|VMTU#LAMBD (55% identity in 132 amino acids)
 SEQ ID NO: 568: -0.34, 251, a putative minor tail protein, similar to minor tail proteins for example ,GpZ
 6545 [Bacteriophage lambda] gi|138849|sp|P03731|VMTZ#LAMBD (52% identity in 206 amino acids)
 SEQ ID NO: 569: -0.141667, 133, novel
 SEQ ID NO: 570: -0.45942, 208, novel (hypothetical membrane protein)
 6550 SEQ ID NO: 571: -0.103226, 94, novel

- SEQ ID NO: 572 : 0.549074, 109, a transposase (OrfA) (insertion sequence IS629), identical to hypothetical protein [Escherichia coli plasmid p O-157 insertion sequence IS629] gi|7444868|pir||T00241
- 6555 SEQ ID NO: 573 : -0.202367, 339, a transposase (OrfB) (insertion sequence IS629), identical to transposase [Escherichia coli plasmid p O-157 insertion sequence IS629] gi|7443862|pir||T00240
- SEQ ID NO: 574 : -0.958333, 109, a putative
- 6560 protease/scaffold protein, similar to ClpP proteases for example ,[Bacteriophage D3] gi|5059251|gb|aaD38956.1| (39% identity in 195 amino acids); putative scaffolding protein [Streptococcus thermophilus bacteriophage DT1] gi|4530143|gb|aaD21883.1| (31% identity in 193 amino acids),
- 6565 GTG start
- SEQ ID NO: 575: -0.397973, 297, a putative portal protein, similar to portal protein-like protein [Wolbachia sp. wKue] gi|6723246|dbj|Baa89642.1| (24% identity in 438 amino acids); similar to(at low level) portal proteins for
- 6570 example ,gp4 [phage 21] gi|549295|sp|P36272|VG04#BPP21 (20% identity in 368 amino acids)
- SEQ ID NO: 576: -0.101359, 369, novel
- SEQ ID NO: 577: -0.4932, 501, a putative terminase large subunit, similar to terminase large subunit-like protein
- 6575 [Wolbachia sp. wKue] gi|6723244|dbj|Baa89640.1| (25% identity in 629 amino acids); terminase large subunits for example ,GpA [Bacteriophage lambda] gi|137616|sp|P03708|TERL#LAMBDA (23% identity in 615 amino acids), GTG start
- 6580 SEQ ID NO: 578: -0.598718, 79, novel
- SEQ ID NO: 579: -0.665488, 708, a lipoprotein Rz1 precursor, similar to lipoproteinRz1 precursors for example ,[Bacteriophage 933W] gi|4585425|gb|aaD25453.1|AF125520#48 (98% identity in 61

6585 amino acids)
 SEQ ID NO: 580: -0.458861, 159, an endopeptidase (host cell lysis), identical to Rz[Bacteriophage VT2-Sa] gi|5881639|dbj|Baa84330.1|; similar to endopeptidases for example ,Rz [Bacteriophage lambda] gi|119368|sp|P00726|ENPP#LAMBD (69% identity in 153 amino acids)
 SEQ ID NO: 581: 0.214754, 62, a putative antirepressor protein, identical to putative antirepressor protein [Bacteriophage 933W] gi|4585423|gb|aaD25451.1|AF125520#46; its N-terminal part (amino acids at the position 1-126) is similar to antirepressor proteinAnt [Bacteriophage P22] (49% identity in 126 amino acids)
 SEQ ID NO: 582: -0.472903, 156, a putative endolysin, identical to endolysin [Bacteriophage 933W] gi|4585422|gb|aaD25450.1|AF125520#45 ; similar to endolysins for example ,R protein [Bacteriophage H-19B] gi|4335686|gb|aaD17382.1| (93% identity in 177 amino acids)
 SEQ ID NO: 583: -0.283069, 190, a putative holin protein, identical to putative holin [Bacteriophage 933W] gi|4499808|emb|CAB39307.1|; similar to holin proteins for example , protein [Bacteriophage 21] gi|138706|sp|P27360|VLYS#BPP21 (77% identity in 71 amino acids)
 SEQ ID NO: 584: -0.449153, 178, novel, similar to hypothetical proteins for example ,[Shigella dysenteriae] gi|6759966|gb|aaF28124.1|AF153317#20 (91% identity in 81 amino acids)
 SEQ ID NO: 585: 0.039437, 72, novel, identical to hypothetical protein [Bacteriophage 933W] gi|4499806|emb|CAB39305.1|
 SEQ ID NO: 586: -0.312346, 82, novel, similar to hypothetical proteins for example ,[Bacteriophage VT2-Sa]

gi|5881634|dbj|Baa84325.1| (92% identity in 649 amino acids)
6620 SEQ ID NO: 587: 0.008475, 60, a Shiga toxin I subunit B precursor, identical to Shiga toxin I subunit B precursor
gi|134539|sp|P08027|SLTB#BPH30
SEQ ID NO: 588: -0.218518, 649, a Shiga toxin I subunit A precursor, identical to Shiga toxin I subunit A precursor
6625 [Shigella dysenteriae] gi|134537|sp|P10149|SLTA#BPH30
SEQ ID NO: 589: 0.031461, 90, an antitermination protein, similar to antitermination proteins for example , protein Q [Bacteriophage H-19B] (95% identity in 144 amino acids)
SEQ ID NO: 590 : 0.083492, 316, novel, similar to
6630 hypothetical proteins for example ,Nin 68 [Bacteriophage lambda] gi|1351593|sp|P03771|Y68#LAMBD (80% identity in 60 amino acids)
SEQ ID NO: 591 : -0.268056, 145, novel, similar to hypothetical proteins for example ,NinG protein
6635 [Bacteriophage 21] gi|4539482|emb|CAB39991.1| (90% identity in 201 amino acids)
SEQ ID NO: 592: -0.534375, 65, novel, similar to hypothetical proteins for example ,NinF [Bacteriophage P22] gi|512350|emb|Caa55162.1| (96% identity in 58 amino acids)
6640 SEQ ID NO: 593: -1.045273, 202, novel
SEQ ID NO: 594 : -0.286957, 70, novel, identical to hypothetical protein [Bacteriophage VT2-Sa] gi|5881625|dbj|Baa84316.1|; similar to Nin E proteins for example ,[Bacteriophage 21] (100% identity in 57 amino acids)
6645 SEQ ID NO: 595 : -0.939098, 134, novel, similar to hypothetical proteins for example ,[Bacteriophage VT2-Sa] gi|5881624|dbj|Baa84315.1| (98% identity in 175 amino acids); DNA N-6-adenine-methyltransferase [Bacteriophage T1] (31% identity in 143 amino acids)
6650 SEQ ID NO: 596: -1.339655, 59, novel, similar to hypothetical proteins for example ,[Bacteriophage 933W] gi|4585410|gb|aaD25438.1|AF125520#33 (98% identity in

148 amino acids); Nin B [Bacteriophage 21]
gi|4539479|emb|CAB39988.1| (43% identity in 147 amino
6655 acids)
SEQ ID NO: 597 : -0.174286, 176, novel, similar to
hypothetical proteins for example ,[Bacteriophage SEQ ID
NO: 933W] gi|4585409|gb|aaD25437.1|AF125520#32 (99%
identity in 109 amino acids), GTG start
6660 SEQ ID NO: 598 : -0.739189, 149, novel, similar to
hypothetical proteins for example ,[Bacteriophage 933W]
gi|4499788|emb|CAB39287.1| (97% identity in 92 amino acids)
SEQ ID NO: 599: 0.00851, 142, a Ren protein, similar to Ren
proteins for example ,[Bacteriophage lambda]
6665 gi|139473|sp|P03761|VREN#LAMB D (97% identity in 96
amino acids)
SEQ ID NO: 600: -0.872826, 93, a phage replication protein P,
similar to phage replication protein Ps for
example ,[Bacteriophage lambda]
6670 gi|139488|sp|P03689|VRPP#LAMB D(97% identity in 233
amino acids)
SEQ ID NO: 601: -0.0375, 97, a phage replication protein O,
similar to phage replication protein Os for
example ,[Bacteriophage 933W]
6675 gi|4585405|gb|aaD25433.1|AF125520#28(99% identity in 312
amino acids)
SEQ ID NO: 602: -0.448927, 234, a regulatory protein CII,
similar to regulatory protein CIIs for
example ,[Bacteriophage 933W]
6680 gi|4585404|gb|aaD25432.1|AF125520#27 (94% identity in 98
amino acids)
SEQ ID NO: 603 : -0.815064, 313, a putative regulatory
protein, similar to putative regulatory proteins for
example ,[Bacteriophage VT2-Sal gi|5881616|dbj|Baa84307.1|
6685 (42% identity in 71 amino acids)
SEQ ID NO: 604 : -0.220408, 99, a putative prophage

repressor CI, similar to prophagerepressor CIs for example [Bacteriophage lambda] gi|133353|sp|P03034|RPC1#LAMBD (48% identity in 205 amino acids)

6690 SEQ ID NO: 605: -0.223611, 73, novel

SEQ ID NO: 606: -0.193868, 213, novel (hypothetical membrane protein)

SEQ ID NO: 607: -0.194624, 94, a putative regulatory protein (transcription anti-termination), similar to putative transcriptionanti-termination proteins for example, protein N [Bacteriophage phi-21] gi|132274|sp|P07243|R for example, N#BPPH3 (99% identity in 64 amino acids)

6695 SEQ ID NO: 608: -0.036066, 184, novel

6700 SEQ ID NO: 609: -0.355556, 91, a putative superinfection exclusion protein, similar to superinfection exclusion protein B [Bacteriophage P22] gi|585991|sp|P38396|SIEB#BPP22 (84% identity in 191 amino acids)

SEQ ID NO: 610: 0.358824, 52, a putative single-stranded DNA binding protein, identical to putative single-stranded DNA binding protein [Bacteriophage 933W] ; similar to Ea10(single-stranded DNA binding protein) [Bacteriophage lambda] gi|137630|sp|P03757|VE10#LAMBD (99% identity in 122 amino acids)

6705 SEQ ID NO: 611: -0.012435, 194, a regulatory protein cIII (antitermination), identical to regulatory protein cIII [Bacteriophage lambda] gi|133366|sp|P03044|RPC3#LAMBD

SEQ ID NO: 612: -0.263935, 123, a Kil protein (host killing), similar to Kil proteins for example, [Bacteriophage lambda] gi|138622|sp|P03758|VKIL#LAMBD (97% identity in 89 amino acids)

6715 SEQ ID NO: 613: -0.544444, 55, a host-nuclease inhibitor protein Gam (interrupted), similar to N-terminal part of gam [Bacteriophage lambda] (99% identity in 37 amino acids)

6720 SEQ ID NO: 614: -0.120225, 90, putative host-nuclease

inhibitor proteinGam, similar to C-terminal part of Gam
[Bacteriophage lambda] gi|138128|sp|P03702|VGAM#LAMBD
(99% identity in 98 amino acids), probably disrupted by
frameshift

6725 SEQ ID NO: 615: -0.28, 51, a recombination protein Bet,
identical to Bet protein[Bacteriophage 933W]
gi|4585391|gb|aaD25419.1|AF125520#14 ; similar to Bet
protein [Bacteriophage lambda]
gi|137511|sp|P03698|VBET#LAMBD (99% identity in 261
6730 amino acids)
SEQ ID NO: 616: -0.707143, 99, an exonuclease, identical to
exonucleases [Bacteriophage933W]
gi|4585390|gb|aaD25418.1|AF125520#13 ; similar to
exonucleases for example ,[Bacteriophage lambda]
6735 gi|119702|sp|P03697|EXO#LAMBD (97% identity in 225 amino
acids)
SEQ ID NO: 617: -0.509195, 262, novel, identical to
hypothetical protein [Bacteriophage 933W]
gi|4585389|gb|aaD25417.1|AF125520#12; similar to
6740 hypothetical protein orf60a [Bacteriophage lambda]
gi|508995|gb|aaA96568.1| (95% identity in 62 amino acids)
SEQ ID NO: 618: -0.358667, 226, novel, identical to
hypothetical protein [Bacteriophage 933W]
gi|4585388|gb|aaD25416.1|AF125520#11; similar to orf63
6745 [Bacteriophage lambda] gi|508994|gb|aaA96567.1| (88%
identity in 61 amino acids)
SEQ ID NO: 619: -0.13871, 63, novel, identical to
hypothetical proteins for example ,[Bacteriophage 933W]
gi|4585387|gb|aaD25415.1|AF125520#10 ; similar to
6750 hypothetical protein orf61 [Bacteriophage lambda] (93%
identity in 46 amino acids)
SEQ ID NO: 620: -0.192064, 64, a putative C4-type zinc
finger protein (TraRfamily), similar to putative C4-type zinc
finger protein (TraR family) for

Appendix B: Hideo *et al.* Full Translation

6755 example ,gi|7649830|dbj|Baa94108.1| (93% identity in 73 amino acids)
SEQ ID NO: 621: -0.410753, 94, novel, its N-terminal part is similar to hypothetical proteins for example ,[Bacteriophage 933W] gi|4585455|gb|aaD25483.1|AF125520#78 (68% identity
6760 in 168 amino acids); its C-terminal part is similar to hypothetical protein [Bacteriophage HK022] gi|6863138|gb|aaF30379.1|AF069308#27 (96% identity in 196 amino acids), GTG start
SEQ ID NO: 622: -0.617808, 74, novel
6765 SEQ ID NO: 623: -0.622222, 316, novel, its N-terminal part (amino acids at the position 1-44) is similar to hypothetical proteins for example ,[Bacteriophage 933W] gi|4585382|gb|aaD25410.1|AF125520#5 (84% identity in 44 amino acids)
6770 SEQ ID NO: 624: -0.068966, 59, novel, partially similar to hypothetical proteins for example ,[Bacteriophage 933W] gi|4585455|gb|aaD25483.1|AF125520#78 (41% identity in 90 amino acids)
SEQ ID NO: 625: -0.482204, 119, novel
6775 SEQ ID NO: 626: -0.8125, 121, a putative excisionase, similar to putative excisionases for example ,[Bacteriophage 933W] gi|4585379|gb|aaD25407.1|AF125520#2 (47% identity in 74 amino acids)
SEQ ID NO: 627: -0.72, 81, a putative integrase, similar to
6780 integrases for example ,[Bacteriophage 933W] gi|4585378|gb|aaD25406.1|AF125520#1 (65% identity in 423 amino acids)
SEQ ID NO: 628 : -0.803572, 85, a putative salicylate hydroxylase, similar to salicylatehydroxylases for
6785 example ,[Streptomyces coelicolor] gi|7481300|pir||T36193 (31% identity in 348 amino acids)
SEQ ID NO: 629 : -0.471028, 429, similar to probable glutathione-S-transferase,glutathione-S-transferases for

example ,[Pseudomonas sp. U2] gi|3406829|gb|aaC29501.1|
6790 (43% identity in 210 amino acids)
SEQ ID NO: 1444 : -0.21864, 398, a putative isomerase,
similar to isomerases for example ,isomerase-decarboxylase
homolog [Pseudomonas sp. U2]
gi|3406828|gb|aaC29500.1|(46% identity in 188 amino acids);
6795 similar to hypothetical protein Orf2 [Sphingomonas sp. RW5]
gi|3550668|emb|Caa12268.1| (54% identity in 228 amino
acids)
SEQ ID NO: 1445 : 0.236279, 216, probable gentisate
1,2-dioxygenase, similar to gentisate 1,2-dioxygenases for
6800 example ,[Pseudomonas alcaligenes]
gi|5733104|gb|aaD49427.1|AF173167#1 (53% identity in 333
amino acids); [Sphingomonas sp. RW5]
gi|3550667|emb|Caa12267.1| (45% identity in 339 amino
acids)
6805 SEQ ID NO: 1446 : -0.183691, 234, a putative transporter
protein, similar to transporter proteins for
example ,4-hydroxybenzoate transporter [Pseudomonas putida]
gi|6093655|sp|Q51955|PCAK#PSEPU (42% identity in 420
amino acids)
6810 SEQ ID NO: 1447 : -0.411988, 343, a putative regulatory
protein, similar to regulatory proteins for example ,galactose
binding protein regulatory element [Azospirillum brasilense]
gi|1730232|sp|P52661|GBPR#AZOBR (32% identity in 281
amino acids)
6815 SEQ ID NO: 1448 : 0.803097, 453, a putative antibiotic
resistance protein, similar to antibiotic resistance protein
homolog YwoG [Bacillus subtilis] gi|7474437|pir|B70065
(38% identity in 381 amino acids)
SEQ ID NO: 1449 : -0.049371, 319, a putative transcription
6820 regulatory element, similar to putative transcription
regulatory elements for example ,YvbU [Bacillus subtilis]
gi|6648030|sp|O32255 (32% identity in 266 amino acids)

SEQ ID NO: - : 0.973737, 397, novel

SEQ ID NO: - : 0.093836, 293, a transposase (OrfA) (insertion
6825 sequence IS629), hypothetical protein
gi|7444868|pir||T00241

SEQ ID NO: 1623, ECs3123:3078013-3079083; -0.672472, 357,
identical to transposase (OrfB) (insertion sequence IS629)
gi|7443862|pir||T00240,

6830 SEQ ID NO: 1653: -0.965741, 109, similar to B2332#ECOLI
gi|7466328|pir||B65006 (41% identity in 289 amino acids)

SEQ ID NO: 1654: -0.397973, 297, similar to B2333#ECOLI
gi|1788674 (56% identity in 174 amino acids); minor fimbrial
subunit StfG protein [Salmonella typhimurium]

6835 dad|AF093503-7|aaC64157.1 (48% identity in 139 amino acids)

SEQ ID N O-1572: -0.075, 281, similar to B2334#ECOLI
gi|1788675 (53% identity in 141 amino acids); similar to minor
fimbrial subunits for example ,StfF [Salmonella typhimurium]
gi|3747033 (53% identity in 158 amino acids)

6840 SEQ ID N O-1573: 0.123626, 183, similar to B2335#ECOLI
gi|1788676 (47% identity in 166 amino acids); similar to minor
fimbrial subunit StfE protein [Salmonella typhimurium]
dad|AF093503-5|aaC64155.1 (48% identity in 154 amino acids)

SEQ ID N O-1574: -0.085256, 157, similar to YFCS#ECOLI
6845 gi|1788677 (85% identity in 250 amino acids); periplasmic
fimbrial chaperone StfD protein [Salmonella typhimurium]
dad|AF093503-4|aaC64154.1 (59% identity in 233 amino acids)

SEQ ID N O-1575: 0.534337, 167, its N-terminal part (amino
acids at the position 1-581/883) is similar to YFCU#ECOLI
6850 gi|1788679 (90% identity in 577 amino acids), its C-terminal
part (amino acids at the position 587-883/883) is similar to
B2337#ECOLI gi|1788678 (88% identity in 297 amino acids)

SEQ ID NO: - : -0.305159, 253, similar to B2339#ECOLI
gi|1788680 (88% identity in 187 amino acids); major fimbrial
6855 subunit StfA protein [Salmonella typhimurium]
dad|AF093503-2|aaC64152.1 (39% identity in 187 amino acids)

SEQ ID NO: - : -0.461661, 880, a putative DNA injection protein, its N-terminal part is similar to N-terminal part of DNA injection protein gp20 [phage P22]
6860 gi|1174950|sp|Q01076|VG20#BPP22(47% identity in 217 amino acids); its C-terminal part is similar to(at low level) C-terminal part of hypothetical proteins for example ,[Caenorhabditis el for example ,ans] gi|5805382|gb|aaD51972.1|AF173372#1 (34% identity in 76
6865 amino acids)
SEQ ID NO: - : -0.20107, 188, a putative DNA transfer protein precursor, similar to DNA transfer protein Gp7 [Bacteriophage P22] gi|418222|sp|Q01074|VG07#BPP22(66% identity in 207 amino acids)
6870 SEQ ID NO: 1289 : -0.056085, 379, novel, similar to hypothetical protein P31 [Bacteriophage APSE-1] gi|6118026|gb|aaF03974.1|AF157835#31 (35% identity in 152 amino acids); gp14 [Bacteriophage P22] gi|418225|sp|Q01075|VG14#BPP22(22% identity in 143 amino
6875 acids)
SEQ ID NO: 1290: -0.180088, 227, novel
SEQ ID NO: 1291: -0.107742, 156, a putative replication protein, partially similar to replication proteins for example ,[Haemophilus actinomycetemcomitans plasmid pVT736-1] gi|398106|gb|aaC37125.1| (26% identity in 145 amino acids)
6880
SEQ ID NO: 1292: 0.176842, 96, novel
SEQ ID NO: 1293: -0.803463, 232, novel
SEQ ID NO: 1294: -1.430769, 53, novel
6885 SEQ ID NO: 1295: -0.364681, 471, a putative resolvase, similar to resolvases for example ,[plasmid pM3] gi|5668998|gb|aaD46124.1|AF078924#3 (46% identity in 204 amino acids); [Yersinia pestis plasmid pMT1] gi|7467461|pir||T14990 (43% identity in 193 amino acids)
6890 SEQ ID NO: 1296: -0.218966, 59, a sucrose transporter protein,

similar to sucrose transporter protein (permease) [Escherichia coli strain EC3132| gi|231914|sp|P30000|CSCB#ECOLI (99% identity in 415 amino acids)

6895 SEQ ID NO: 1297: -0.367308, 209, a putative fructokinase, similar to fructokinase (EC 2.7.1.4) for example, [Escherichia coli strain EC3132| gi|730731|sp|P40713|SCRK#ECOLI (98% identity in 291 amino acids)

6900 SEQ ID NO: 1298: 0.823615, 416, a sucrose hydrolase, similar to sucrose hydrolase [Escherichia coli strain EC3132| gi|3462879|gb|aaC33123.1| (98% identity in 477 amino acids)

SEQ ID NO: 1299: 0.010855, 305, a sucrose operon repressor, sucrose operon repressor [Escherichia coli]

SEQ ID NO: similar to gi|729214|sp|P40715|CSCR#ECOLI (99% identity in 331 amino acids)

6905 SEQ ID NO: 1300: -0.532914, 478, similar to EryA homologue [Bacteriophage If1| dad|U02303-9|aaC62159.1 (76% identity in 333 amino acids)

SEQ ID NO: 1301: -0.041088, 332, a putative transposase, similar to transposase homologA [Helicobacter pylori]

6910 gi|2114470|gb|aaD11513.1 (58% identity in 137 amino acids)

SEQ ID NO: 1618: -0.604712, 383, similar to FLXA#ECOLI gi|2498386|sp|P77609 (43% identity in 74 amino acids)

SEQ ID NO: - : -0.437222, 181, a putative polyferredoxin, similar to ferredoxin [Methanosarcina thermophila]

6915 gi|282643|pir|A42960 (48% identity in 43 amino acids); similar to polyferredoxin [Methanococcus voltae] gi|99156|pir|S24802 (22% identity in 207 amino acids)

SEQ ID NO: - : -0.478761, 114, a putative anaerobic dimethyl sulfoxide reductase chain C, similar to anaerobic dimethyl

6920 sulfoxide reductase chain Cs for example, [Escherichia coli] gi|118699|sp|P18777|DMSC#ECOLI (27% identity in 271 amino acids)

SEQ ID NO: 1490: -0.1, 285, a putative anaerobic dimethyl sulfoxide reductasechain B, similar to anaerobic dimethyl

6925 sulfoxide reductases chain Bs for example ,[Escherichia coli]
gi|2506394|sp|P18776|DMSB#ECOLI (59% identity in 185
amino acids)
SEQ ID NO: 1491 : 1.152381, 274, a putative anaerobic
dimethyl sulfoxide reductase chain A precursor, similar to
6930 anaerobic dimethyl sulfoxide reductase chain A precursors for
example ,[Escherichia coli]
gi|118697|sp|P18775|DMSA#ECOLI (43% identity in 768
amino acids)
SEQ ID NO: 1492 : -0.325837, 210, novel, similar to DNA
6935 damage-inducible proteins for example ,DinI [Escherichia coli]
gi|2498305|sp|Q47143|DINI#ECOLI (43% identity in 81 amino
acids)
SEQ ID NO: 1493 : -0.412988, 794, novel, similar to(at low
level) putative Cys3His zinc finger protein ATCTH
6940 [Arabidopsis thaliana] gi|1800279|gb|aaB68046.1| (37%
identity in 35 amino acids)
SEQ ID NO: 1061 : -0.60122, 83, a chaperone-like protein,
similar to TrcA-like proteins for example ,bfpT-r for
example ,ulated chaperone-like protein TrcA [Escherichiacoli
6945 strain B171-8] gi|4126789|dbj|Baa36747.1| (85% identity in
195 amino acids)
SEQ ID NO: 1062 : -0.528302, 54, novel, similar to
hypothetical proteins for example ,ORF2 [Escherichia coli
strain B171-8] gi|4126790|dbj|Baa36748.1| (99% identity in
6950 216 amino acids)
SEQ ID NO: 1063 : -0.526531, 197, novel, similar to
hypothetical protein ORF3[Escherichia coli strain B171-8]
gi|4126791|dbj|Baa36749.1| (98% identity in 352 amino acids)
SEQ ID NO: 1064 : -0.181019, 217, novel, similar to
6955 hypothetical proteins for example ,ORF4 [Escherichia coli
strain B171-8] gi|4126792|dbj|Baa36750.1| (99% identity in
140 amino acids)
SEQ ID NO: 1065 : -0.571307, 353, novel, similar to

Appendix B: Hideo *et al.* Full Translation

hypothetical protein [Bacteriophage 933W]
6960 gi|4585437|gb|aaD25465.1|AF125520#60 (93% identity in 129 amino acids)
SEQ ID NO: 1066: -0.416429, 141, identical to transposase, hypothetical protein [Escherichia coli plasmid p O-157 insertion sequence IS629] gi|7444868|pir||T00241; similar to
6965 hypothetical protein, IS elements for example ,TnpE[Shigella flexneri] gi|5532454|gb|aaD44738.1|AF141323#9 (97% identity in 108 amino acids)
SEQ ID NO: 1067: -0.251938, 130, a transposase, identical to transposase [Escherichiacoli plasmid p O-157 insertion
6970 sequence IS629] gi|7443862|pir||T00240
SEQ ID NO: 1068: -0.965741, 109, novel, its N-terminal part (amino acids at the position 1-87) is partially similar to hypothetical proteins for example ,L0015 (amino acids at the position 50-136/512) [Escherichia coli O-157:H7 strain EDL933]
6975 gi|3414883|gb|aaC31494.1|
SEQ ID NO: 1069: -0.397973, 297, novel, identical to hypothetical protein L0014[Escherichia coli O-157:H7 strain EDL933] gi|3288157|emb|Caa11510.1|; similar to hypothetical proteins for example ,ORF50 [Escherichia coli]
6980 gi|6009426|dbj|Baa84885.1| (76% identity in 107 amino acids)
SEQ ID NO: 1070: -0.501818, 166, novel, similar to hypothetical proteins for example ,L0013 [Escherichia coli O-157:H7 strain EDL933] gi|3414881|gb|aaC31492.1| (100% identity in 126 amino acids)
6985 SEQ ID NO: 1071: 0.010435, 116, a putative endolysin (host cell lysis), similar to N-terminal-half part of endolysins for example ,[Bacteriophage 933W]
gi|4585422|gb|aaD25450.1|AF125520#45 (93% identity in 73 amino acids), probably interrupted
6990 SEQ ID NO: 1072: -0.403175, 127, novel, similar to hypothetical protein YdfR[Escherichia coli] gi|3183262|sp|P76160|YDFR#ECOLI (47% identity in 74

Appendix B: Hideo *et al.* Full Translation

amino acids)
 SEQ ID NO: 1073: -0.144737, 77, a holin (host cell lysis),
 6995 similar to holin proteins for example ,[Bacteriophage VT2-Sa]
 gi|5881636|dbj|Baa84327.1| (90% identity in 91 amino acids)
 SEQ ID NO: 1074: -0.027193, 115, novel, similar to
 hypothetical proteins for example ,[Bacteriophage 933W]
 gi|4585419|gb|aaD25447.1|AF125520#42 (52% identity in 613
 7000 amino acids)
 SEQ ID NO: 1075: 0.095775, 72, novel
 SEQ ID NO: 1076: -0.210048, 618, novel, similar to
 hypothetical proteins for example ,[Actinobacillus
 actinomycescomitans] gi|7592819|dbj|Baa94406.1 (29%
 7005 identity in 228 amino acids)
 SEQ ID NO: 1077: 0.446789, 110, antitermination, similar to
 antitermination proteins for example , protein Q
 [Bacteriophage lambda] gi|132278|sp|P03047|R for
 example ,Q#LAMBD (97% identity in 207 amino acids)
 7010 SEQ ID NO: 1078: 0.628745, 248, a serine/threonine protein
 phosphatase, similar to serine/threonine proteinphosphatases
 for example ,[Bacteriophage lambda]
 gi|130792|sp|P03772|PP#LAMBD (95% identity in 221 amino
 acids)
 7015 SEQ ID NO: 1079: -0.263768, 208, novel, similar to
 hypothetical proteins for example ,NinG [Bacteriophage 21]
 gi|4539482|emb|CAB39991.1| (89% identity in 199 amino
 acids)
 SEQ ID NO: 1080: -0.243891, 222, novel, similar to phage
 7020 hypothetical proteins for example ,[Bacteriophage
 phi-YeO3-12] gi|6598993|emb|CAB63597.1| (32% identity in
 110 amino acids)
 SEQ ID NO: 1081: -1.078325, 204, a putative transposase,
 similar to N-terminal part of transposases for
 7025 example ,[Escherichia coli insertion sequence IS30]
 gi|2851554|sp|P37246|TRA8#ECOLI (100% identity in 247

- amino acids)
- SEQ ID NO: 1082: -0.772872, 189, novel, TTG start
- SEQ ID NO: 1083: -0.849402, 252, novel
- 7030 SEQ ID NO: 1084: -0.28168, 132, novel
- SEQ ID NO: 1085: -1.133413, 423, novel
- SEQ ID NO: 1086: -0.535766, 138, novel, its C-terminal part is similar to ctp synthase - *Rickettsia prowasekii* gi|7438005|pir||C71695 (24% identity in 138 amino acids); its
- 7035 N-terminal part is similar to hypothetical protein - *Plasmodium falciparum* gi|4493974|emb|CAB39033.1| (24% identity in 129 amino acids)
- SEQ ID NO: 1087: -0.442424, 133, novel
- SEQ ID NO: 1088: -0.501657, 544, a putative integrase,
- 7040 similar to site specific recombinases for example ,integraserecombinase protein [*Methanobacterium thermoautotrophicum*] gi|7428936|pir||D69219 (27% identity in 174 amino acids)
- SEQ ID NO: 1089: -0.314416, 438, novel (DNA-binding protein), similar to putative DNA-binding protein [*Bacteriophage P4*] gi|140147|sp|P12552|Y9K#BPP4 (42% identity in 50 amino acids); similar to hypothetical proteins for example ,[*Yersinia pestis*] gi|7467337|pir||T17447 (46% identity in 40 amino acids)
- 7045
- 7050 SEQ ID NO: 1090: -0.426185, 402, novel
- SEQ ID NO: 1091: -0.441176, 69, a putative regulatory element, similar to regulatory proteins for example ,MocR [*Sinorhizobium meliloti*] gi|1346565|sp|P49309 (34% identity in 466 amino acids)
- 7055
- 7055 SEQ ID NO: 1092: -0.333569, 284, novel, similar to conserved hypothetical protein [*Streptomyces coelicolor* A3(2)] gi|7649565|emb|CAB89054.1 (38% identity in 141 amino acids)
- SEQ ID NO: 1597: -0.168469, 445, novel, similar to N-terminal part of hypothetical proteins for example ,VdcD
- 7060 [*Streptomyces* sp. D7] gi|4741970|gb|aaD28783.1|AF134589#3

(57% identity in 71 amino acids); YclD [Bacillus subtilis] gi|7452267|pir|A69762 (48% identity in 68 amino acids)

SEQ ID NO: 1598 : -0.074126, 144, a putative 4-hydroxybenzoate decarboxylase, identical to YclC

7065 [Escherichia coli O-157:H7 strain?] gi|4887556|emb|CAB43499.1| (100% identity in 475 amino acids); similar to VdcC [Streptomyces sp. D7] gi|6686069|sp|Q9X697|VDCC#STRD7 (72% identity in 474 amino acids); 4-hydroxybenzoate decarboxylase [Clostridium

7070 hydroxybenzoicum] gi|5739200|gb|aaD50377.1|AF128880#1(53% identity in 469 amino acids)

SEQ ID NO: 1541: -0.65, 79, a putative phenylacrylic acid decarboxylase, identical to Pad1[Escherichia coli O-157:H7

7075 strain ?] gi|4887557|emb|CAB43500.1|; similar to phenylacrylic acid decarboxylases for example ,VdcB [Streptomyces sp. D7] (73% identity in 190 amino acids)

SEQ ID NO: 1542: -0.214105, 476, a transcription regulatory element, identical to SlyA [Escherichia coli O-157:H7 strain ?]

7080 to gi|4887558|emb|CAB43501.1|; similar to transcription regulatory elements for example ,[Streptomyces coelicolor] gi|7481485|pir|T35022 (32% identity in 124 amino acids)

SEQ ID NO: 1543 : 0.027919, 198, novel, similar to hypothetical proteins for example ,[Escherichia coli]

7085 gi|7404494|sp|P45956|YGBF#ECOLI (86% identity in 94 amino acids)

SEQ ID NO: 1544 : -0.374074, 136, novel, similar to hypothetical protein b2755[Escherichia coli strain K-12] gi|7460139|pir|G65056 (84% identity in 303 amino acids),

7090 GTG start

SEQ ID NO: 1330 : 0.025773, 98, novel, similar to(at low level) hypothetical protein b2756 [Escherichia coli strain K-12] gi|6136707|sp|Q46897|YGCH#ECOLI (28% identity in 200 amino acids)

- 7095 SEQ ID NO: 1331 : -0.038111, 308, novel, similar to hypothetical protein b2757[Escherichia coli strain K-12] gi|7459357|pir| |A65057 (35% identity in 160 amino acids)
SEQ ID NO: 1332 : -0.411111, 217, novel, similar to hypothetical protein b2758[Escherichia coli strain K-12] gi|7476186|pir| |C70849 (32% identity in 93 amino acids) [0022]
- 7100 5) Regulatory element
Sequence number: hydrophobicity, The number of amino acids, Character such as function
- 7105 SEQ ID NO: 1333: -0.537097, 249, novel
SEQ ID NO: 1334 : -0.248718, 352, novel, similar to hypothetical protein b2760[Escherichia coli strain K-12] gi|7451979|pir| |D65057 (24% identity in 303 amino acids)
SEQ ID NO: 1335 : -0.612921, 179, novel, similar to hypothetical protein YgcB[Escherichia coli strain K-12] gi|2506493|sp|P38036|YGCB#ECOLI (28% identity in 778 amino acids), GTG start
SEQ ID NO: 1336: -0.429615, 521, similar to YBDY#ECOLI gi|3025009|sp|P77091 (78% identity in 50 amino acids);
- 7115 similar to SrnB [plasmid F] dad|AP001918-5|Baa97875.1 (42% identity in 49 amino acids)
SEQ ID NO: 1337 : -0.257627, 886, novel, similar to hypothetical proteins for example ,Tp70 [Treponema pallidum] gi|7521576|pir| |A71309 (35% identity in 124 amino acids)
- 7120 SEQ ID NO: - : 0.81, 51, novel, similar to N-terminal part of hypothetical proteins for example ,YgcG [Escherichia coli] gi|1723817|sp|P55140|YGCG#ECOLI(43% identity in 186 amino acids)
- 7125 SEQ ID NO: 1512: -0.608397, 132, novel
SEQ ID NO: 1513: 0.301786, 225, novel, its N-terminal part is similar to N-terminal part of hypothetical proteins for example ,YgcG [Escherichia coli]

gi|1723817|sp|P55140|YGCG#ECOLI(31% identity in 147
7130 amino acids)
SEQ ID NO: 1514 : 0.238, 51, similar to YGCG#ECOLI
gi|1789140 (40% identity in 275 amino acids); similar to
hypothetical protein [Pseudomonas aeruginosa]
dad|AE004490-5|aaG03925.1 (43% identity in 273 amino acids),
7135 GTG start
SEQ ID NO: 1515: 0.225393, 383, a lipoprotein precursor (type
III secretion system), similar to type III secretion system
lipoprotein precursors for example ,PrgK protein [Salmonella
typhimurium] gi|1172615|sp|P41786|PRGK#SALTY (53%
7140 identity in 231 amino acids)
SEQ ID NO: : 0.151648, 274, a type III secretion protein,
similar to MxiI [Shigella flexneri]
gi|547954|sp|Q06080|MXII#SHIFL (32% identity in 93 amino
acids);PrgJ protein [Salmonella typhimurium]
7145 gi|1172614|sp|P41785|PRGJ#SALT Y (31% identity in 87
amino acids)
SEQ ID NO: 1192: 0.037705, 245, a type III secretion protein,
similar to putative typeIII secretion proteins for
example ,PrgI protein [Salmonella
7150 typhimurium]gi|1172613|sp|P41784|PRGI#SALTY (64%
identity in 76 amino acids)
SEQ ID NO: 1193: -0.282727, 111, a putative adherence factor,
similar to a part of adherence factors for example ,Efa1
[Escherichia coli O111:H- strain E45035]
7155 gi|6013469|gb|aaD49229.2|AF159462#1(amino acids at the
position 433-711/3223) (100% identity in 279 amino acids),
probably disrupted by frameshift
SEQ ID NO: 1194: -0.588608, 80, a transposase, identical to
transposase [Escherichia coli plasmid p O-157 IS629]
7160 gi|7443862|pir|T00240
SEQ ID NO: 1195: -0.379918, 245, a transposase, identical to
hypothetical protein [Escherichia coli plasmid p O-157

IS629] gi|7444868|pir||T00241; similar to hypothetical protein, insertion sequences for example ,[Shigella flexneri]

7165 gi|5532454|gb|aaD44738.1|AF141323#9 (96% identity in 108 amino acids)

SEQ ID NO: 1196: -0.045181, 167, novel, GTG start

SEQ ID NO: 1197 : -0.081233, 374, novel, similar to hypothetical proteins for example ,L0014 [Escherichia coli

7170 O-157:H7 strain EDL933] gi|3414882|gb|aaC31493.1| (99% identity in 115 amino acids)

SEQ ID NO: 1198 : 1.038462, 79, novel, similar to hypothetical proteins for example ,L0015 [Escherichia coli

7175 O-157:H7 strain EDL933] gi|3414883|gb|aaC31494.1|(100% identity in 411 amino acids)

SEQ ID NO: 1199: 0.805162, 151, novel, similar to a part of hypothetical proteins for example ,L0013 [Escherichia coli

O-157:H7 strain EDL933] gi|3414881|gb|aaC31492.1| (55% identity in 28 amino acids), GTG start, probably disrupted

7180 SEQ ID NO: 1200 : 0.976744, 87, novel, similar to hypothetical proteins for example ,ORF50 [Escherichia coli plasmid pB171] gi|6009426|dbj|Baa84885.1| (70% identity in 106 amino acids)

SEQ ID NO: 1201 : 0.748416, 222, novel, similar to

7185 hypothetical proteins for example ,L0015 [Escherichia coli O-157:H7 strain EDL933] gi|3414883|gb|aaC31494.1|(63% identity in 464 amino acids)

SEQ ID NO: 1202: -0.236585, 329, novel, similar to a part of transposases for example ,TnpA [Shigella flexneri]

7190 gi|5532449|gb|aaD44733.1|AF141323#4 (93% identity in 49 amino acids)

SEQ ID NO: 1203 : -1.506341, 206, novel, similar to hypothetical proteins for example ,L0004 [Escherichia coli

7195 O-157:H7 strain EDL933] gi|3414872|gb|aaC31483.1| (98% identity in 91 amino acids); putative transposase [Vibrio cholerae] gi|7960026|gb|aaF71186.1|AF179596#6 (59%

- identity in 91 amino acids); hypothetical protein [Escherichia coli plasmid p O-157 insertion sequence IS911] gi|7465897|pir||T00224 (52% identity in 91 amino acids)
- 7200 SEQ ID NO: 1204: -0.892208, 78, a putative transcription regulatory element, similar to regulatory elements (RpiR family) for example ,[Bacillus subtilis] gi|8248807|emb|CAB93068.1| (25% identity in 236 amino acids)
- 7205 SEQ ID NO: 1205 : -1.002703, 112, a putative ferrichrome-binding protein, similar to ferrichrome-binding proteins for example ,[Bacillus subtilis] gi|585132|sp|P37580|FHUD#BACSU (27% identity in 220 amino acids)
- 7210 SEQ ID NO: 1206: -0.212558, 440, a putative ferrichrome ABC transporter (permease), similar to ferrichrome ABC transporters (permease) for example ,[Bacillus subtilis] gi|1706797|sp|P49937|FHUG#BACSU (33% identity in 319 amino acids)
- 7215 SEQ ID NO: 1207: 0.465452, 687, a putative ferrichrome ABC transporter (permease), similar to ferrichrome ABC transporters (permease) for example ,[Synechocystis sp.] gi|7442493|pir||S74438 (43% identity in 315 amino acids); [Bacillus subtilis] gi|1706795|sp|P49936|FHUB#BACSU (39% identity in 319 amino acids)
- 7220 SEQ ID NO: 1208: -0.209449, 382, a putative ABC-type iron-siderophore transport system ATP-binding protein, similar to ABC-type iron-siderophore transport system ATP-binding proteins for example ,[Synechocystis sp.] gi|7442509|pir||S74440 (52% identity in 248 amino acids)
- 7225 SEQ ID NO: 1209: -0.149383, 568, a putative ferrichrome-iron receptor precursor, similar to ferrichrome-iron receptor precursors for example ,gi|7448497|pir||S74457 (30% identity in 688 amino acids)
- 7230 SEQ ID NO: 1210: 0.036546, 250, novel, TTG start

SEQ ID NO: 1211 : 1.166101, 60, a PTSdependent N-acetyl-galactosamine-IID component (AgaE), similar to PTSdependent N-acetyl-galactosamine-IID component, AgaE [Escherichia coli strain C]

7235 gi|8895749|gb|aaF81085.1|AF228498#5 (96% identity in 292 amino acids)

SEQ ID NO: - : -0.257895, 77, a PTS dependent N-acetyl-galactosamine-and galactosamine IIA component (AfaF), similar to ts dependent N-acetyl-galactosamine-and galactosamine IIA component, AgaF [Escherichia coli strain C]

7240 gi|8895750|gb|aaF81086.1|AF228498#6 (99% identity in 144 amino acids)

SEQ ID NO: 1527 : 0.06993, 144, a transposase (insertion sequence IS629), identical to hypothetical protein

7245 gi|7444868|pir||T00241

SEQ ID NO: 1528 : 1.167709, 193, identical to transposase (insertion sequence IS629),gi|7443862|pir||T00240

SEQ ID NO: 1529 : 0.38766, 236, novel

SEQ ID NO: 1530 : -0.008, 226, a leader peptidase, similar to leader peptidases for example ,HopD (strain ECOR30) [Escherichia coli] gi|7674073|sp|O68932 (92% identity in 155 amino acids); (LT2) [Salmonella typhimurium] gi|7674072|sp|O68927 (68% identity in 148 amino acids)

7250

SEQ ID NO: 1531 : -0.168, 226, novel, similar to hypothetical protein [Xylellafastidiosa]

7255 gi|9112262|gb|aaF85593.1|AE003851#24 (50% identity in 86 amino acids)

SEQ ID NO: - : -0.265401, 238, a putative invasins, similar to putative membrane protein b1978 [Escherichia coli K-12]

7260 gi|1736642|dbj|Baa15799.1| (45% identity in 1391 amino acids); vasin [Yersinia pseudotuberculosis] gi|79202|pir||A29646 (35% identity in 1211 amino acids) [0023]

6) Proteins relating to fimbriae

- 7265 Sequence number: hydrophobicity, The number of amino acids,
Character such as function
 SEQ ID NO: 1674 : -0.352675, 244, similar to replication
 protein O, for example , protein O [Enterobacteria phage
 HK022| gi|407289|gb|aaB60272.1| (98% identity in 299 amino
 7270 acids)
 SEQ ID NO: 1129 : -0.391449, 422, a replication protein P
 (putative replication DNAhelicase), similar to P proteins
 for example ,[Enterobacteria phage HK022|
 gi|6863143|gb|aaF30384.1|AF069308#32 (99% identity in 478
 7275 amino acids); replication DNA helicases for example ,DnaB
 [Escherichia coli| gi|118713|sp|P03005|DNAB#ECOLI (39%
 identity in 436 amino acids)
 SEQ ID NO: 1130 : -0.275728, 207, novel, identical to
 hypothetical protein [Bacteriophage VT2-Sa]
 7280 gi|5881620|dbj|Baa84311.1| (100% identity in 89 amino acids)
 SEQ ID NO: 1131 : -0.090099, 102, novel, identical to
 hypothetical protein [Bacteriophage 933W]
 gi|4499788|emb|CAB39287.1| (100% identity in 92 amino
 acids)
 7285 SEQ ID NO: 1132 : -0.513839, 225, a type III secretion protein,
 similar to PrgH protein [Salmonella typhimurium]
 gi|1172612|sp|P41783|PRGH#SALTY (28% identity in 266
 amino acids); MxiG [Shigella flexneri]
 gi|2498603|sp|Q57332|MXIG#SHIFL (23% identity in 243
 7290 amino acids)
 SEQ ID NO: 1133 : -0.08, 116, a putative transcription
 regulatory element, similar to transcription activator
 NtrC[Herbaspirillum seropedicae]
 gi|57313501|gb|aaC32391.21 (25% identity in 107 amino acids)
 7295 SEQ ID NO: 1134 : -0.503734, 483, a type III secretion protein,
 similar to type IIIsecretion proteins for example ,SpaS protein
 [Salmonella typhimurium] gi|730801|sp|P40702|SPAS#SALTY
 (54% identity in 348 amino acids)

- SEQ ID NO: 1135: -0.293631, 315, novel,
- 7300 SEQ ID NO: 1136: -0.452748, 183, ABC transporter (binding protein), similar to binding proteins for example ,phosphate-binding protein PstS homolog [Methanobacterium thermoautotrophicum (strain Delta H)] gi|7442891|pir| |A69098 (32% identity in 187 amino acids)
- 7305 SEQ ID NO: 1137: 0.39434, 54, its N-terminal part (amino acids at the position 1-77/505) is similar to YZGL#ECOLIgi|1789834 (83% identity in 77 amino acids); its C-terminal part (amino acids at the position 325-519/525) is similar to binding proteins for example ,phosphate-binding
- 7310 protein PstS homolog [Methanobacterium thermoautotrophicum strain Delta H] gi|7442891|pir| |A69098 (31% identity in 175 amino acids)
- SEQ ID NO: 1138: 0.390909, 67, a putative DNA processing chain A, similar to many DNA processing chain As (Smf
- 7315 protein), for example ,[Neisseria meningitidis] gi|7378929|emb|CAB83472.1| (30% identity in 265 amino acids)
- SEQ ID NO: 1139: -0.774999, 297, a putative ATP-dependent DNA helicase (partial), similar to C-terminal part of
- 7320 ATP-dependent DNA helicase [Streptomyces coelicolor] gi|7480492|pir| |T35189(64% identity in 37 amino acids), GTG start
- SEQ ID NO: 1140: -0.122667, 76, a putative ATP-dependent DNA helicase (partial), similar to a part of ATP-dependent
- 7325 DNA helicase [Streptomyces coelicolor] gi|7480492|pir| |T35189 (31% identity in 269 amino acids), GTG start
- SEQ ID NO: 1141: -0.286338, 550, a putative ATP-dependent DNA helicase (partial), similar to a part of putative
- 7330 ATP-dependent DNA helicase [Streptomyces coelicolor] gi|7480492|pir| |T35189 (48% identity in 175 amino acids)
- SEQ ID NO: 1142: -0.02069, 59, a putative ATP-dependent

DNA helicase (interrupted), similar to N-terminal part of putative ATP-dependent DNA helicases for example ,[*Streptomyces coelicolor*] gi|7428315|pir||T35189 (60% identity in 176 amino acids); [*Bacillus subtilis*] gi|7436435|pir||F69901 (42% identity in 169 amino acids)

7335 SEQ ID NO: 1143: -0.395745, 330, novel
SEQ ID NO: 1144 : -0.477678, 225, novel (hypothetical membrane protein)

7340 SEQ ID NO: 1145 : -0.43168, 263, novel (hypothetical membrane protein)
SEQ ID NO: 1146 : -0.74642, 434, novel, similar to hypothetical protein ORF79[*Escherichia coli* plasmid pB171] gi|6009455|dbj|Baa84914.1 (62% identity in 175 amino acids)

7345 SEQ ID NO: 1147 : -0.610909, 276, novel, similar to hypothetical protein ORF80[*Escherichia coli* plasmid pB171] (70% identity in 86 amino acids)

7350 SEQ ID NO: 1148 : -0.397973, 297, novel (hypothetical lipoprotein)
SEQ ID NO: 1149 : -0.965741, 109, a putative O-methyltransferase, similar to a part of O-methyltransferases for example ,acetylserotonin N-methyltransferase (EC 2.1.1.4) -

7355 chicken gi|2498445|sp|Q92056|HIOM#CHICK (28% identity in 157 amino acids)
SEQ ID NO: 1150: -0.836842, 39, novel
SEQ ID NO: 1151: 0.029565, 116, a putative acyltransferase, similar to acyltransferases for example ,[*Neisseria meningitidis* MC58] gi|7226953|gb|aaF42046.1| (33% identity in 246 amino acids)

7360 SEQ ID NO: 1152: -0.409503, 464, a putative acyl carrier protein, similar to acyl carrier proteins for example ,[*Neisseria meningitidis* MC58] gi|7226952|gb|aaF42045.1| (51% identity in 85 amino acids)

7365 SEQ ID NO: 1153: -0.178846, 53, a putative acyl carrier

protein, similar to acyl carrier proteins for example ,[*Neisseria meningitidis* MC58] gi|7226951|gb|aaF42044.1| (51% identity in 79 amino acids)

7370 SEQ ID NO: 1154 : -0.063793, 117, novel (hypothetical membrane protein), similar to putative integral membrane protein [*Neisseria meningitidis*] gi|7380586|emb|CAB85174.1| (51% identity in 126 amino acids)

7375 SEQ ID NO: 1155 : -0.55546, 468, novel, similar to peptide synthetase [sic, synthase] [*Xylella fastidiosa*] gi|9105980|gb|aaF83848.1|AE003941#2 (26% identity in 420 amino acids); p-coumaryl-CoA ligase [*Rhodobacter sphaeroides*] gi|2764724|emb|Caa05380.1| a part of (27% identity in 268 amino acids); a part of surfactin synthetase component I [7380 [*Bacillus subtilis*] gi|2127235|pir||I40485 (20% identity in 410 amino acids)]

7385 SEQ ID NO: 1156 : -0.569643, 57, a putative (3R)-hydroxymyristoyl-[acyl carrier protein] dehydratase, similar to(at low level) a part of (3R)-hydroxymyristoyl-[acyl carrier protein] dehydratases for example ,[*Salmonella typhimurium*] gi|140182|sp|P21773|FABZ#SALTY (29% identity in 67 amino acids)

7390 SEQ ID NO: - : -0.908772, 115, novel, its N-terminal part is similar to dolichyl-phosphate mannose synthase related proteins for example ,[*Pyrococcus abyssi* (strain Orsay)] gi|7445533|pir||A75176 (30% identity in 206 amino acids); its N-terminal part is similar to HmsR [*Yersinia pestis*] gi|1185391|gb|aaB66590.1| (34% identity in 128 amino acids); its C-terminal part is similar to hypothetical protein [*Xylella fastidiosa*] gi|9105669|gb|aaF83585.1|AE003918#7 (30% identity in 310 amino acids)

7395 SEQ ID NO: 1402 : 0.001017, 296, novel, similar to hypothetical proteins for example ,.[*Deinococcus radiodurans*] gi|7471367|pir||B75463 (31% identity in 111 amino acids), GTG start

7400

Appendix B: Hideo *et al.* Full Translation

SEQ ID NO: 1403: -0.013016, 316, novel
 SEQ ID NO: 1404: 1.044986, 350, novel, similar to membrane
 protein [Xylella
 fastidiosa|gi|9105671|gb|aaF83587.1|AE003918#9 (24%
 7405 identity in 502 amino acids)
 SEQ ID NO: 1405: 1.132416, 328, novel
 SEQ ID NO: 1406 : -0.004833, 270, putative
 3-oxoacyl-(acyl-carrier protein) synthase II, similar to
 3-oxoacyl-(acyl-carrier protein) synthase IIs for
 7410 example ,[Streptomyces coelicolor A3(2)]
 gi|7479090|pir|T34912 (31% identity in 381 amino acids)
 SEQ ID NO: 1407 : -0.402244, 714, a putative
 beta-hydroxydecanoyl-ACP dehydrase, similar to hypothetical
 protein [Neisseria meningitidis MC58]
 7415 gi|7226956|gb|aaF42049.1| (32% identity in 116 amino acids);
 beta-hydroxydecanoyl-ACP dehydrase [Pseudomonas
 aeruginosa|gi|2384563|gb|aaC45619.1| (29% identity in 123
 amino acids)
 SEQ ID NO: - : -0.405385, 131, a putative
 7420 3-oxoacyl-(acyl-carrier protein) reductase, similar to
 3-oxoacyl-(acyl-carrier protein) reductases for
 example ,[Neisseria meningitidis MC58]
 gi|7226957|gb|aaF42050.1| (57% identity in 242 amino acids)
 SEQ ID NO: 1585 : 0.50548, 293, similar to putative
 7425 3-oxoacyl-(acyl-carrier protein)synthase IIs for
 example ,gi|7226958|gb|aaF42051.1| (48% identity in 404
 amino acids)
 SEQ ID NO: 1586: 0.152083, 145, a putative transcription
 regulatory element, similar to transcription regulatory
 7430 elements for example ,[Escherichia coli]
 gi|129347|sp|P13669|FARR#ECOLI (28% identity in 235
 amino acids)
 SEQ ID NO: 1656 : -0.965741, 109, a putative PTS
 (phosphotransferase system) system enzyme IIA, similar to PTS

- 7435 system enzyme IIA components for example ,[*Escherichia coli* K-12] gi|2507274|sp|P37187|PTKA#ECOLI (23% identity in 122 amino acids): PTSsystem fructose-specific enzyme IIBC component [*Bacillus halodurans*] gi|4512375|dbj|Baa75339.1| (33% identity in 151 amino acids)
- 7440 SEQ ID NO: 1657: -0.397973, 297, a putative PTS system enzyme IIB, similar to PTS system, galactitol-specific IIB component [*Escherichia coli* K-12] gi|2507273|sp|P37188|PTKB#ECOLI (35% identity in 92 amino acids)
- 7445 SEQ ID NO: - : 0.072131, 62, a putative PTS system enzyme IIC, similar to PTS system galactitol-specific enzyme IICs for example ,[*Bacillus halodurans*] gi|4512376|dbj|Baa75340.1| (45% identity in 411 amino acids)
- SEQ ID NO: 1695: 0.74129, 156, a putative sugar kinase, similar to sugar kinases for example ,xylulokinase (EC 2.7.1.17) [*Lactobacillus pentosus*] gi|139850|sp|P21939|XYLB#LACPE (23% identity in 496 amino acids)
- 7450
- SEQ ID NO: 1678: -0.385107, 95, a putative PTS system HPr enzyme, similar to phosphotransferase system HPr enzymes for example ,[*Xylella fastidiosa*] gi|9106413|gb|aaF84212.1|AE003971#11 (39% identity in 87 amino acids)
- 7455
- SEQ ID NO: 1679: 0.150932, 162, a putative aldolase, similar to aldolases for example ,[*Vibrio furnissii*] gi|1732204|gb|aaC44684.1| (38% identity in 272 amino acids)
- 7460
- SEQ ID NO: - : 0.763317, 200, novel, similar to HicB-related protein [*Xylella fastidiosa*] gi|9106728|gb|aaF84477.1|AE003992#13 (35% identity in 110 amino acids); HicB [*Haemophilus influenzae*] gi|3603326|gb|aaC35810.1| (26% identity in 93 amino acids)
- 7465
- SEQ ID NO: 1548: -0.459394, 331, novel, similar to HicA [*Haemophilus influenzae*] gi|3603325|gb|aaC35809.1| (30%

identity in 60 amino acids)

7470 [0024]

7) Proteins relating to transportation of substance
Sequence number: hydrophobicity, The number of amino acids,
Character such as function

SEQ ID NO: - : 0.123763, 506, a type III secretion protein,
 7475 similar to C-terminal part of type III secretion proteins for
 example ,SpaR protein [Salmonella typhimurium]
 gi|730799|sp|P40701|SPAR#SALTY(56% identity in 65 amino
 acids), may be partial (disrupted by frameshift)

SEQ ID NO: 1521 : -0.08725, 401, novel, similar to
 7480 hypothetical protein [Xylella fastidiosa]
 gi|9112263|gb|aaF85594.1|AE003851#25 (48% identity in 158
 amino acids)

SEQ ID NO: 1522: 0.754902, 52, novel

SEQ ID NO: 1523: -0.310185, 325, heme utilization/transporter
 7485 protein, identical to ChuA [Escherichia coli O-157:H7 EDL933]
 gi|1763009|gb|aaC44857.1|

SEQ ID NO: 1524: 0.080682, 177, novel, TTG start

SEQ ID NO: 1525: -0.081683, 203, a putative hemin-binding
 protein, similar to hypothetical protein huT [Shigella
 7490 dysenteriae haem transport locus] gi|2967538|gb|aaC27815.1|
 (97% identity in 304 amino acids); hemin-binding proteins for
 example ,[Yersinia pestis]
 gi|6226635|sp|Q56991|HMUT#YERPE (34% identity in 253
 amino acids)

7495 SEQ ID NO: 1613 : -0.262046, 304, a putative
 coproporphyrinogen oxidase, similar to coproporphyrinogen
 oxidases for example ,PhuW [Vibrio parahaemolyticus
 gi|5106980|gb|aaD39908.1|AF119047#1 (35% identity in 371
 amino acids)

7500 SEQ ID NO: 1614 : 0.671015, 139, novel, similar to
 hypothetical proteinhuX [Shigella dysenteriae haem transport
 locus] gi|2967537|gb|aaC27814.1| (98% identity in 164 amino

acids); hypothetical protein X [Yersinia pestis]
gi|7467368|pir|T12066 (60% identity in 153 amino acids)

7505 SEQ ID NO: 1659 : -0.222178, 249, novel, similar to
hypothetical proteinhuY [Shigella dysenteriae haem transport
locus] gi|2967536|gb|aaC27813.1| (97% identity in 207 amino
acids); hypothetical protein Y [Yersinia pestis]
gi|7467369|pir|T12067 (55% identity in 204 amino acids)

7510 SEQ ID NO: - : -0.069143, 176, a putative hemin permease,
similar to hypothetical proteinhuU [Shigella dysenteriae haem
transport locus] gi|2967535|gb|aaC27812.1| (99% identity in
318 amino acids); hemin permeases for example ,HmuU
[Yersinia pestis] gi|6226636|sp|Q56992|HUU#YERPE (66%
7515 identity in 318 amino acids)
SEQ ID NO: 1671: -0.626137, 89, a putative hemin transport
system ATP-binding protein, similar to hypothetical
proteinhuV [Shigella dysenteriae haem transport locus]
gi|2967534|gb|aaC27811.1| (98% identity in 256 amino acids);

7520 hemin transport systemATP-binding proteins for
example ,HmuV [Yersinia pestis]
gi|2492539|sp|Q56993|HMUV#YERPE(58% identity in 264
amino acids)
SEQ ID NO: 1241: -0.4456, 126, a putative fimbrial protein
7525 precursor, similar to fimbrial proteins for example ,long polar
fimbrial minor protein precursor [Salmonellatyphimurium]
gi|1170819|sp|P43664|LPFE#SALTY (50% identity in 165
amino acids)
SEQ ID NO: 1242 : 0.022946, 354, a putative fimbrial
7530 protein precursor, similar to fimbrial proteins for
example ,long polar fimbrial protein LpFD [Salmonella
typhimurium] gi|1170818|sp|P43663|LPFD#SALTY (39%
identity in 350 amino acids)
SEQ ID NO: 1243 : -0.201546, 195, a putative outer
7535 membrane usher proteinLpFC precursor (partial), similar to C
-terminal-half part of outer membrane usher proteins for

- example ,LpfC precursor [Salmonella typhimurium]
gi|1170817|sp|P43662|LPFC#SALTY(67% identity in 485
amino acids), GTG start
- 7540 SEQ ID NO: 1244: 0.154275, 270, a putative outer membrane
usher protein, similar to N-terminal-half part of outer
membrane usher proteins for example ,LpfC [Salmonella
typhimurium] gi|1170817|sp|P43662|LPFC#SALTY (69%
identity in 357 amino acids), interrupted TAG stop codon
- 7545 SEQ ID NO: 1245: 0.251765, 86, a putative fimbrial
chaperone protein, similar to chaperones for example ,LpfB
[Salmonella typhimurium]
gi|1170816|sp|P43661|LPFB#SALTY (67% identity in 229
amino acids)
- 7550 SEQ ID NO: 1246: -0.375904, 84, a putative fimbrial major
protein precursor, similar to long polar fimbria proteinA
precursor, LpfA, of S. typhimurium,
gi|1170815|sp|P43660|LPFA#SALTY (73% identity in 178
amino acids)
- 7555 SEQ ID NO: 1247: 0.721244, 194, a putative transcription
regulatory element, similar to(at low level)hypothetical
transcription regulator yisR [Bacillus subtilis]
gi|3123306|sp|P40331 (24% identity in 276 amino acids)
- 7560 SEQ ID NO: 1248: -0.13819, 454, a putative permease,
similar to hypothetical protein [Salmonella typhimurium]
gi|7442781|pir|C65167 (37% identity in 444 amino acids);
transporter proteins (putative symporters) for example ,YicJ
[Escherichia coli (K-12)] gi|2851421|sp|P31435|YICJ#ECOLI
(32% identity in 340 amino acids)
- 7565 SEQ ID NO: 1249: -0.388034, 118, novel, similar to
hypothetical protein [Thermotoga maritima]
gi|7452109|pir|F72395 (37% identity in 635 amino acids)
- SEQ ID NO: 1250: -0.070968, 559, novel, similar to
hypothetical protein [Neisseria meningitidis MC58]
7570 gi|7227012|gb|aaF42100.1 (39% identity in 398 amino acids)

SEQ ID NO: 1251: -0.387143, 141, novel, TTG start
 SEQ ID NO: 1252: -0.435323, 202, novel, TTG start
 SEQ ID NO: 1253: 0.383311, 750, novel, similar to surface
 proteins, for example ,[Xylella fastidiosa]
 7575 gi|9106565|gb|aaF84338.1|AE003982#11 (24% identity in 1514
 amino acids)
 SEQ ID NO: 1254 : -0.125258, 195, identical to lipid
 A-core:surface polymer ligase (WaaL), WaaL [Escherichia coli
 strain F653] gi|3821825|gb|aaC69661.1| (100% identity in 402
 7580 amino acids)
 SEQ ID NO: 1255: -0.00874, 390, similar to lipopolysaccharide
 1,2-N acetylglucosaminetransferase (WaaD), WaaD [Escherichia
 coli strain F653] gi|3821826|gb|aaC69662.1| (99% identity in
 380 amino acids)
 7585 SEQ ID NO: 1256 : 0.065584, 155, a putative
 UDP-glucose:(galactosyl) LPS alpha1, similar to
 2-glucosyltransferase (WaaJ), UDP-glucose:(galactosyl) LPS
 alpha1,2-glucosyltransferase WaaJ [Escherichia coli strain
 F653] gi|3821827|gb|aaC69663.1| (98% identity in 184 amino
 7590 acids), TTG start
 SEQ ID NO: 1257: 0.147325, 244, a lipopolysaccharide core
 biosynthesis, identical to WaaY [Escherichia coli strain F653]
 gi|3821828|gb|aaC69664.1| (100% identity in 235 amino acids)
 SEQ ID NO: - : -0.156479, 410,
 7595 UDP-D-galactose:(glucosyl)lipopolysaccharide-
 alpha-1,3-D-galactosyltransferase, similar to WaaI (strain F653
 R3 core type)[Escherichia coli] gi|3821829|gb|aaC69665.1
 (99% identity in 335 amino acids)
 SEQ ID NO: 1427: -0.248606, 252, novel
 7600 SEQ ID NO: 1428 : 0.024841, 158, a putative integrase,
 identical to CP4-like integrase [Escherichia coli EDL933]
 gi|3414871|gb|aaC31482.1|; similar to integrases for
 example ,[Shigella flexneri]
 gi|5532446|gb|aaD44730.1|AF141323#1 (95% identity in 390

- 7605 amino acids)
 SEQ ID NO: 1429: 0.37957, 94, novel, identical to L0004
 [Escherichia coli strain EDL933] gi|3414872|gb|aaC31483.1|;
 similar to hypothetical proteins for example ,[Escherichia
 coli plasmid p O-157 insertion sequence IS911
 7610 gi|7465897|pir|T00224 (56% identity in 116 amino acids),
 GTG start
 SEQ ID NO: 1430: 0.897123, 453, novel, identical to L0005
 [Escherichia coli strain EDL933] gi|3414873|gb|aaC31484.1|,
 GTG start
 7615 SEQ ID NO: 1431: -0.065339, 503, novel, identical to L0006
 [Escherichia coli strain EDL933] gi|3414874|gb|aaC31485.1|;
 similar to hypothetical proteins for example ,[Vibrio
 cholerae] gi|7960027|gb|aaF71187.1|AF179596#7 (60%
 identity in 300 amino acids)
 7620 SEQ ID NO: 1432: -0.496629, 90, novel, similar to C-terminal
 part of hypothetical proteins for example ,b2004 (YeeU)
 [Escherichia coli] gi|3025157|sp|P76364|YEEU#ECOLI(84%
 identity in 53 amino acids)
 SEQ ID NO: 1433: -0.054196, 287, novel, identical to L0007
 7625 [Escherichia coli EDL933]gi|3414875|gb|aaC31486.1|; similar
 to hypothetical proteins for example ,b2005(yeeV)
 [Escherichia coli] gi|3025158|sp|P76365|YEEV#ECOLI (88%
 identity in 124 amino acids)
 SEQ ID NO: 1434: -0.327731, 120, novel, identical to L0008
 7630 [Escherichia coli EDL933]gi|3414876|gb|aaC31487.1|; similar
 to hypothetical protein [Escherichiacoli D1114, O25:K10:H16]
 gi|4887094|gb|aaD32187.1| (90% identity in 114 amino acids);
 similar to b2006 (YeeW) [Escherichia coli]
 gi|3025160|sp|P76366|YEEW#ECOLI (70% identity in 55
 7635 amino acids)
 SEQ ID NO: 1435: -0.472528, 92, novel, identical to L0009
 [Escherichia coli strain EDL933] gi|3414877|gb|aaC31488.1|;
 similar to hypothetical protein [Escherichia coli D1114,

O25:K10:H16] gi|4887094|gb|aaD32187.1|(84% identity in 59
7640 amino acids); hypothetical protein [Salmonella typhi]
gi|7800330|gb|aaF69926.1|AF250878#87 (46% identity in 49
amino acids)
SEQ ID NO: 1339: -0.276608, 343, novel, identical to L0010
[Escherichia coli strain EDL933] gi|3414878|gb|aaC31489.1|;
7645 similar to PH01 [Escherichia coli D1114, O25:K10:H16]
gi|4887092|gb|aaD32185.1|AF127177#3 (62% identity in 78
amino acids)
SEQ ID NO: 1340: -0.474091, 661, novel, similar to(at low
level) a part of hypothetical protein ydiA[plasmid ColIb-P9]
7650 gi|4512489|dbj|Baa75138.1|(42% identity in 35 amino acids)
SEQ ID NO: 1341: -0.667647, 69, novel, identical to L0012
[Escherichia coli EDL933] gi|3414880|gb|aaC31491.1|; similar
to a part of putative ATP-binding proteinugR
[Salmonellatyphimurium] gi|4324607|gb|aaD16951.1| (45%
7655 identity in 66 amino acids)
SEQ ID NO: 1342: 0.113158, 305, novel, identical to L0013
[Escherichia coli EDL933] gi|3414881|gb|aaC31492.1|; similar
to hypothetical proteins for example ,Hp3 [Escherichia coli
CFT073] gi|3661484|gb|aaC61715.1| (100% identity in 74
7660 amino acids)
SEQ ID NO: 1343: -0.308539, 446, novel, identical to L0014
[Escherichia coli] gi|3414882|gb|aaC31493.1|; similar to
hypothetical proteins for example ,orf50 [Escherichia coli
plasmid pB171] gi|6009426|dbj|Baa84885.1| (76% identity in
7665 107 amino acids)
SEQ ID NO: 1344: -0.137195, 165, novel, similar to L0015
[Escherichia coli EDL933]gi|3414883|gb|aaC31494.1| (99%
identity in 512 amino acids); hypothetical proteins for
example ,[Escherichia coli plasmid pEAF]
7670 gi|4808945|gb|aaD30027.1|AF119170#2 (91% identity in 447
amino acids)
SEQ ID NO: 1345: 0.057488, 208, novel, similar to a part of

IS630 insertion element hypothetical protein
gi|1143207|gb|aaA84873.1| (72% identity in 25 amino acids)

7675 SEQ ID NO: 1346: 0.933648, 319, novel, similar to a part of
hypothetical proteins for example ,[insertion sequence lS91]
gi|7466597|pir|T00311 (75% identity in 49 amino acids)
SEQ ID NO: 1347: -0.269531, 257, a secreted effector protein,
identical to L0016 [Escherichia coli EDL933]

7680 gi|3414884|gb|aaC31495.1|; similar to EspF [Escherichia coli
E2348/69] gi|2865308|gb|aaC38400.1| (87% identity in 205
amino acids)
SEQ ID NO: 1461: -0.092614, 177, novel, identical to L0017
[Escherichia coli EDL933]gi|3414885|gb|aaC31496.1|; similar

7685 to hypothetical proteins for example ,[Escherichia coli]
gi|2809428|gb|aaC28566.1| (97% identity in 92 amino acids)
SEQ ID NO: 1462: -0.045584, 352, novel, identical to EscF
[Escherichia coli] gi|2865306|gb|aaC38398.1|; L0018
[Escherichia coli EDL933] gi|3414886|gb|aaC31497.1|

7690 SEQ ID NO: 1463: -0.460825, 486, novel, identical to L0019
[Escherichia coli EDL933]gi|3414887|gb|aaC31498.1|; similar
to hypothetical proteins for example ,Orf27[Escherichia coli
E2348/69] gi|2865305|gb|aaC38397.1| (99% identity in 135
amino acids)

7695 SEQ ID NO: 1464: -0.264578, 368, an EspB protein (secreted
protein), similar to EspB proteins for example ,EspB(L0020)
[Escherichia coli EDL933] gi|1657263|emb|Caa65654.1| (99%
identity in 312 amino acids)
SEQ ID NO: 1465: -0.234061, 230, an EspD secreted protein,

7700 identical to L0021 [Escherichia coli EDL933]
gi|3414889|gb|aaC31500.1|; similar to EspD proteins for
example ,gi|3688279|emb|Caa76909.1| (85% identity in 374
amino acids)
SEQ ID NO: 1466: 0.12809, 179, an EspA secreted protein,

7705 identical to EspA protein (L0022) [Escherichia coli]
gi|3115184|emb|Caa73506.1|; similar to EspA proteins for

example .gi|2388623|gb|aaB71083.1| (85% identity in 192 amino acids)

7710 SEQ ID NO: - : -0.31476, 272, a type III secretion system SepL protein, identical to SepL (L0023) [Escherichia coli EDL933] gi|3115183|emb|Caa73505.1|; similar to SepL proteins for example .gi|2865301|gb|aaC38393.1| (94% identity in 351 amino acids)

7715 SEQ ID NO: 1507: 0.694205, 467, a type III secretion system EscD protein, identical to Pas (L0024) [Escherichia coli EDL933] gi|3115182|emb|Caa73504.1|; similar to EscD proteins for example .gi|3341420|emb|Caa74170.1| (97% identity in 6 amino acids)

7720 SEQ ID NO: - : -0.414177, 657, a Gamma intimin, identical to Gamma intimin (L0025) [Escherichia coli strain EDL933] gi|3414893|gb|aaC31504.1|

7725 SEQ ID NO: - : -0.310441, 432, a chaperon of Tir, identical to CesT [Escherichia coli O-157:H7 strain HA1] gi|975876|gb|aaB00110.1|; similar to CesT protein [Escherichia coli] gi|140611|sp|P21244|YEA#ECOLI (96% identity in 156 amino acids)

7730 SEQ ID NO: - : -0.190991, 112, a translocated intimin receptor Tir, identical to translocated intimin receptor Tir (L0027) [Escherichia coli O-157:H7 strain EDL933] gi|3414895|gb|aaC31506.1| [0025]

S) Proteins relating to synthesis of lipopolysaccharide

Sequence number: hydrophobicity, The number of amino acids, Character such as function

7735 SEQ ID NO: 1333: -0.537097, 249, novel

SEQ ID NO: 1334 : -0.248718, 352, novel, similar to hypothetical protein b2760[Escherichia coli strain K-12] gi|7451979|pir|D65057 (24% identity in 303 amino acids)

7740 SEQ ID NO: 1335 : -0.612921, 179, novel, similar to hypothetical protein YgcB[Escherichia coli strain K-12]

Appendix B: Hideo *et al.* Full Translation

gi|2506493|sp|P38036|YGCB#ECOLI (28% identity in 778 amino acids), GTG start

SEQ ID NO: 1336: -0.429615, 521, similar to YBDY#ECOLI
 7745 gi|3025009|sp|P77091 (78% identity in 50 amino acids);
 similar to SrnB [plasmid F] dad|AP001918-5|Baa97875.1
 (42% identity in 49 amino acids)

SEQ ID NO: 1337: -0.257627, 886, novel, similar to
 hypothetical proteins for example ,Tp70 [Treponema
 pallidum] gi|7521576|pir||A71309 (35% identity in 124 amino
 7750 acids)

SEQ ID NO: - : 0.81, 51, novel, similar to N-terminal part of
 hypothetical proteins for example ,YgcG [Escherichia coli|
 gi|1723817|sp|P55140|YGCG#ECOLI(43% identity in 186
 amino acids)

7755 SEQ ID NO: 1512: -0.608397, 132, novel

SEQ ID NO: 1513: 0.301786, 225, novel, its N-terminal part
 is similar to N-terminal part of hypothetical proteins for
 example ,YgcG [Escherichia coli|
 gi|1723817|sp|P55140|YGCG#ECOLI(31% identity in 147
 7760 amino acids)

SEQ ID NO: 1514: 0.238, 51, similar to YGCG#ECOLI
 gi|1789140 (40% identity in 275 amino acids); similar to
 hypothetical protein [Pseudomonas aeruginosa|
 dad|AE004490-5|aaG03925.1 (43% identity in 273 amino acids),
 7765 GTG start

SEQ ID NO: 1515: 0.225393, 383, a lipoprotein precursor (type
 III secretion system), similar to type III secretion system
 lipoprotein precursors for example ,PrgK protein [Salmonella
 typhimurium] gi|1172615|sp|P41786|PRGK#SALTY (53%
 7770 identity in 231 amino acids)

SEQ ID NO: - : 0.151648, 274, a type III secretion protein,
 similar to MxiI [Shigella flexneri|
 gi|547954|sp|Q06080|MXII#SHIFL (32% identity in 93 amino
 acids);PrgJ protein [Salmonella typhimurium|

Appendix B: Hideo *et al.* Full Translation

7775 gi|1172614|sp|P41785|PRGJ#SALT Y (31% identity in 87 amino acids)
 SEQ ID NO: 1192: 0.037705, 245, a type III secretion protein, similar to putative typeIII secretion proteins for example ,PrgI protein [Salmonella typhimurium]gi|1172613|sp|P41784|PRGI#SALT Y (64% identity in 76 amino acids)
 SEQ ID NO: 1193: -0.282727, 111, a putative adherence factor, similar to a part of adherence factors for example ,Efa1 [Escherichia coli O111:H- strain E45035]
 7785 gi|6013469|gb|aaD49229.2|AF159462#1(amino acids at the position 433-711/3223) (100% identity in 279 amino acids), probably disrupted by frameshift
 SEQ ID NO: 1194: -0.588608, 80, a transposase, identical to transposase [Escherichia coli plasmid p O-157 IS629]
 7790 gi|7443862|pir||T00240
 SEQ ID NO: 1195: -0.379918, 245, a transposase, identical to hypothetical protein [Escherichia coli plasmid p O-157 IS629] gi|7444868|pir||T00241; similar to hypothetical protein, insertion sequences for example ,[Shigella flexneri]
 7795 gi|5532454|gb|aaD44738.1|AF141323#9 (96% identity in 108 amino acids)
 SEQ ID NO: 1196: -0.045181, 167, novel, GTG start
 SEQ ID NO: 1197 : -0.081233, 374, novel, similar to hypothetical proteins for example ,L0014 [Escherichia coli O-157:H7 strain EDL933] gi|3414882|gb|aaC31493.1| (99% identity in 115 amino acids)
 7800 SEQ ID NO: 1198 : 1.038462, 79, novel, similar to hypothetical proteins for example ,L0015 [Escherichia coli O-157:H7 strain EDL933] gi|3414883|gb|aaC31494.1|(100% identity in 411 amino acids)
 7805 SEQ ID NO: 1199: 0.805162, 151, novel, similar to a part of hypothetical proteins for example ,L0013 [Escherichia coli O-157:H7 strain EDL933] gi|3414881|gb|aaC31492.1| (55%

- identity in 28 amino acids), GTG start, probably disrupted
- 7810 SEQ ID NO: 1200 : 0.976744, 87, novel, similar to hypothetical proteins for example ,ORF50 [Escherichia coli plasmid pB171] gi|6009426|dbj|Baa84885.1| (70% identity in 106 amino acids)
- 7815 SEQ ID NO: 1201 : 0.748416, 222, novel, similar to hypothetical proteins for example ,L0015 [Escherichia coli O-157:H7 strain EDL933] gi|3414883|gb|aaC31494.1|(63% identity in 464 amino acids)
- 7820 SEQ ID NO: 1202: -0.236585, 329, novel, similar to a part of transposases for example ,TnpA [Shigella flexneri] gi|5532449|gb|aaD44733.1|AF141323#4 (93% identity in 49 amino acids)
- 7825 SEQ ID NO: 1203 : -1.506341, 206, novel, similar to hypothetical proteins for example ,L0004 [Escherichia coli O-157:H7 strain EDL933] gi|3414872|gb|aaC31483.1| (98% identity in 91 amino acids); putative transposase [Vibrio cholerae] gi|7960026|gb|aaF71186.1|AF179596#6 (59% identity in 91 amino acids); hypothetical protein [Escherichia coli plasmid p O-157 insertion sequence IS911] gi|7465897|pir|T00224 (52% identity in 91 amino acids)
- 7830 SEQ ID NO: 1204: -0.892208, 78, a putative transcription regulatory element, similar to regulatory elements (RpiR family) for example ,[Bacillus subtilis] gi|8248807|emb|CAB93068.1| (25% identity in 236 amino acids)
- 7835 SEQ ID NO: 1205 : -1.002703, 112, a putative ferrichrome-binding protein, similar to ferrichrome-binding proteins for example ,[Bacillus subtilis] gi|585132|sp|P37580|FHUDB#BACSU (27% identity in 220 amino acids)
- 7840 SEQ ID NO: 1206: -0.212558, 440, a putative ferrichrome ABC transporter (permease), similar to ferrichrome ABC transporters (permease) for example ,[Bacillus subtilis]

Appendix B: Hideo *et al.* Full Translation

gi|1706797|sp|P49937|FHUG#BACSU (33% identity in 319 amino acids)

7845 SEQ ID NO: 1207: 0.465452, 687, a putative ferrichrome ABC transporter (permease), similar to ferrichrome ABC transporters (permease) for example ,[Synechocystis sp.] gi|7442493|pir||S74438 (43% identity in 315 amino acids); [Bacillus subtilis] gi|1706795|sp|P49936|FHUB#BACSU (39% identity in 319 amino acids)

7850 SEQ ID NO: 1208: -0.209449, 382, a putative ABC-type iron-siderophore transport system ATP-binding protein, similar to ABC-type iron-siderophore transport system ATP-binding proteins for example ,[Synechocystis sp.] gi|7442509|pir||S74440 (52% identity in 248 amino acids)

7855 SEQ ID NO: 1209: -0.149383, 568, a putative ferrichrome-iron receptor precursor, similar to ferrichrome-ironreceptor precursors for example ,gi|7448497|pir||S74457 (30% identity in 688 amino acids)

7860 SEQ ID NO: 1210: 0.036546, 250, novel, TTG start
SEQ ID NO: 1211: 1.166101, 60, a PTSdependent N-acetyl-galactosamine-IIID component (AgaE), similar to PTSdependent N-acetyl-galactosamine-IIID component, AgaE [Escherichia coli strain C]

7865 gi|8895749|gb|aaF81085.1|AF228498#5 (96% identity in 292 amino acids)
SEQ ID NO: : -0.257895, 77, a PTS dependent N-acetyl-galactosamine-and galactosamine IIA component (AfaF), similar to ts dependent N-acetyl-galactosamine-and galactosamine IIA component, AgaF [Escherichia coli strain C]

7870 gi|8895750|gb|aaF81086.1|AF228498#6 (99% identity in 144 amino acids)
SEQ ID NO: 1527: 0.06993, 144, a transposase (insertion sequence IS629), identical to hypothetical protein

7875 gi|7444868|pir||T00241
SEQ ID NO: 1528: 1.167709, 193, identical to transposase

(insertion sequence IS629),gi|7443862|pir||T00240
 SEQ ID NO: 1529: 0.38766, 236, novel
 SEQ ID NO: 1530: -0.008, 226, a leader peptidase, similar to
 7880 leader peptidases for example ,HopD (strain ECOR30)
 [Escherichia coli] gi|7674073|sp|O68932 (92% identity in 155
 amino acids); (LT2) [Salmonella typhimurium]
 gi|7674072|sp|O68927 (68% identity in 148 amino acids)
 SEQ ID NO: 1531: -0.168, 226, novel, similar to hypothetical
 7885 protein [Xylella fastidiosa]
 gi|9112262|gb|aaF85593.1|AE003851#24 (50% identity in 86
 amino acids)
 SEQ ID NO: - : -0.265401, 238, a putative invasin, similar to
 putative membrane protein b1978 [Escherichia coli K-12]
 7890 gi|1736642|dbj|Baa15799.1| (45% identity in 1391 amino
 acids); vasin [Yersinia pseudotuberculosis]
 gi|79202|pir||A29646 (35% identity in 1211 amino acids)
 [0026]
 9) Proteins relating to metabolism
 7895 Sequence number: hydrophobicity, The number of amino
 acids, Character such as function
 SEQ ID NO: 826: -0.36383, 48, novel, similar to hypothetical
 protein[Bacteriophage 933W] gi|4499789|emb|CAB39288.1|
 (97% identity in 71 amino acids)
 7900 SEQ ID NO: 827: -0.877049, 62, a putative fimbrial chaperone,
 similar to fimbrial chaperones for example ,[Salmonella
 typhimurium] gi|1170816|sp|P43661|LPFB#SALTY (40%
 identity in 104 amino acids)
 SEQ ID NO: 828: -0.388722, 134, a putative type 1 fimbrial
 7905 protein, similar to type 1fimbrial proteins for
 example ,[Salmonella enteritidis] gi|913907|gb|aaB33536.1|
 (31% identity in 198 amino acids)
 SEQ ID NO: 829: 0.010435, 116, novel, similar to conserved
 hypothetical proteins for example ,HP0709 [Helicobacter
 7910 pylori 26695] gi|7463979|pir||E64608 (88% identity in 300

amino acids)
 SEQ ID NO: 830 : -0.455859, 513, novel, similar to
 hypothetical protein [Xylella fastidiosa]
 gi|9104946|gb|aaF82968.1|AE003869#5 (33% identity in 270
 7915 amino acids)
 SEQ ID NO: 831 : -0.335065, 78, novel (hypothetical
 membrane protein)
 SEQ ID NO: 832 : -1.205882, 52, novel, similar to (at low
 level) membrane protein [Staphylococcus aureus]
 7920 gi|3676428|gb|aaC61946.1 (26% identity in 236 amino acids)
 SEQ ID NO: 833: -0.434677, 249, novel
 SEQ ID NO: 834: 0.071739, 93, novel, GTG start
 SEQ ID NO: 835: -0.190411, 74, novel, GTG start
 SEQ ID NO: 836 : -0.322222, 136, a raffinose metabolism
 7925 (putativ for example ,lyco protein), similar to RafY [Escherichia
 coli plasmid pRSD2] gi|1773072|gb|aaB71432.1 (78%
 identity in 464 amino acids)
 SEQ ID NO: 837: -0.195833, 313, novel
 SEQ ID NO: 838 : -0.038235, 375, novel (hypothetical
 7930 membrane protein)
 SEQ ID NO: 839: -0.158854, 193, a Rhs protein, similar to Rhs
 proteins for example ,RhsF[Escherichia coli]
 gi|2920637|gb|aaC32473.1| (97% identity in 1394 amino acids),
 [RhsH core protein with extension]
 7935 SEQ ID NO: 840: -0.174074, 352, novel
 SEQ ID NO: 841 : -0.092611, 407, a putative amino acid
 amidohydrolase, similar to amino acid amidohydrolases for
 example ,benzoylglycine amidohydrolase (Hippuricase)
 [Campylobacter jejuni] gi|1170277|spP45493|HIPO#CAMJE
 7940 (46% identity in 383 amino acids)
 SEQ ID NO: 842 : -0.384796, 935, a putative
 membranetransporter protein, similar to
 membranetransporter proteins for example ,citrate-proton
 symporter [Klebsiella pneumoniae]

- 7945 gi|116482|sp|P16482|CIT1#KLEPN (30% identity in 429 amino acids)
 SEQ ID NO: 843 : -0.174359, 157, novel, similar to hypothetical protein b3122[*Escherichia coli* (strain K-12)]
 gi|7466507|pir|G65101 (62% identity in 35 amino acids)
- 7950 SEQ ID NO: 844 : -0.563799, 559, a putative L-sorbose 1-phosphate dehydrogenase, similar to L-sorbose 1-phosphate dehydrogenases, for example ,[*Klebsiella pneumoniae*]
 gi|586014|sp|P37084|SORE#KLEPN (85% identity in 407 amino acids)
- 7955 SEQ ID NO: 845: -0.552709, 204, a putative sorbose-permease IID component (PTS system), similar to many sorbose-permease IID components for example ,gi|548634|sp|P37083|PTRD#KLEPN (95% identity in 215 amino acids), probably disrupted (N-terminal part (amino acids at the position 1-60) is deleted)
- 7960 SEQ ID NO: 846 : -0.058268, 128, a putative regulatory element (repressor), its N-terminal-half part is similar to hypothetical protein HI1476 [*Haemophilus influenzae*]
 gi|1175815|sp|P44207|YE76#HAEIN (35% identity in 70 amino acids); its C -terminal-half part is similar to putative repressor protein [Bacteriophage D108]
 gi|133345|sp|P07040|RPC1#BPD10(26% identity in 79 amino acids)
- 7970 SEQ ID NO: 847: -0.457738, 169, a putative DNA-binding protein, similar to Ner-like DNA-binding proteins for example ,gi|6900348|emb|CAB71960.1| (44% identity in 70 amino acids)
- 7975 SEQ ID NO: 848 : -0.023279, 306, a putative phage transposase, similar to transposases for example ,[*Neisseria meningitidis*]
 gi|7379960|emb|CAB84536.1| (40% identity in 639 amino acids)
- SEQ ID NO: 849 : -0.484058, 139, a transposition protein, similar to DNA transposition proteinB [Bacteriophage Mu]

gi|139318|sp|P03763|VPB#BPMU (48% identity in 298 amino
7980 acids)
SEQ ID NO: 850: -0.9296, 126, novel, similar to(at low level)
phosphoserine phosphatase [Neisseria meningitidis MC58]
gi|7226221|gb|aaF41385.1| (38% identity in 49 amino acids)
SEQ ID NO: 851: 0.013677, 447, novel
7985 SEQ ID NO: 852: 0.371556, 676, novel, GTG start
SEQ ID NO: 853: 0.247863, 118, novel, GTG start
SEQ ID NO: 854: 0.445454, 100, novel
SEQ ID NO: 855: -0.008451, 143, putative host-nuclease
inhibitor, similar to host-nuclease inhibitor protein (Gam) for
7990 example ,[Bacteriophage Mu]
gi|138127|sp|P06023|VGAM#BPMU (56% identity in 174
amino acids)
SEQ ID NO: 856: -0.096842, 191, novel
SEQ ID NO: 857: -0.295364, 152, novel, similar to Gp11
7995 [Bacteriophage Mu] gi|6010385|gb|aaF01088.1|AF083977#7
(67% identity in 177 amino acids)
SEQ ID NO: 858: -0.149414, 513, novel, similar to gp12
[Bacteriophage Mu] gi|215568|gb|aaA32400.1| (52% identity
in 168 amino acids)
8000 SEQ ID NO: 859: -0.454967, 152, novel, similar to gp9
[Bacteriophage Mu] gi|6010430|gb|aaF01133.1|AF083977#54
(30% identity in 82 amino acids)
SEQ ID NO: 860: -0.538686, 138, novel
SEQ ID NO: 861: -0.001626, 124, novel, similar to (at low
8005 level) zinc finger proteins for example ,[Rattus norvegicus]
gi|141712|sp|P18745|ZO22#XENLA (33% identity in 48 amino
acids)
SEQ ID NO: 862: -0.207895, 153, novel
SEQ ID NO: 863: 0.275652, 346, novel, similar to
8010 hypothetical proteins for example ,gp16 [Bacteriophage Mu]
gi|6010390|gb|aaF01093.1|AF083977#12 (43% identity in 162
amino acids)

SEQ ID NO: 864: 1.013566, 259, putative positive regulator of late transcription, similar to transcription regulatory elements for example , positive regulator of late transcription (protein C) [Bacteriophage Mu] gi|139320|sp|P06022|VPC#BPMU (39% identity in 126 amino acids)

8015

SEQ ID NO: 865: 1.206742, 90, an endolysin (host cell lysis), similar to endolysins for example ,Lys [Bacteriophage Mu] |126600|sp|P27359|LYCV#BPP21 (37% identity in 156 amino acids)

8020

SEQ ID NO: 866 : 0.813365, 218, novel, similar to P14 [Bacteriophage APSE-1] gi|6118009|gb|aaF03957.1|AF157835#14 (27% identity in 82 amino acids), GTG start

8025

SEQ ID NO: 867 : -0.361905, 232, novel, similar to P16 [Bacteriophage APSE-1] gi|6118011|gb|aaF03959.1|AF157835#16 (46% identity in 81 amino acids)

8030

SEQ ID NO: 868: -0.288945, 200, novel, similar to traR family, for example ,Orf82 [Bacteriophage P2] gi|732223|sp|Q06424|YO82#BPP2 (52% identity in 34 amino acids)

8035

SEQ ID NO: 869: -0.829907, 108, novel, similar to gp25 [Bacteriophage Mu] gi|6010400|gb|aaF01103.1|AF083977#22 (35% identity in 91 amino acids)

SEQ ID NO: 870: -0.475, 73, novel, similar to hypothetical proteins for example ,gp26[Bacteriophage Mu] gi|6010401|gb|aaF01104.1|AF083977#23 (62% identity in 95 amino acids)

8040

SEQ ID NO: 871 : -0.715504, 130, novel, similar to hypothetical proteins for example ,gp27 [Bacteriophage Mu] gi|6010402|gb|aaF01105.1|AF083977#24 (56% identity in 193 amino acids)

8045

SEQ ID NO: 872: 0.351219, 42, a putative portal protein,

similar to hypothetical proteins for example ,gp28 (possible portal protein H) [Bacteriophage Mu] gi|6010403|gb|aaF01106.1|AF083977#25 (73% identity in 537 amino acids)

8050 SEQ ID NO: 873 : -0.262814, 399, novel, similar to hypothetical proteins for example ,gp29 [Bacteriophage Mu] gi|6010404|gb|aaF01107.1|AF083977#26 (57% identity in 529 amino acids)

8055 SEQ ID NO: 874 : -0.127574, 273, novel, similar to hypothetical proteins for example ,gp30 [Bacteriophage Mu] gi|6010405|gb|aaF01108.1|AF083977#27 (60% identity in 437 amino acids)

8060 SEQ ID NO: 875 : -0.857143, 78, a virion morphogenesis, similar to G protein [Bacteriophage Mu] gi|267389|sp|Q01261|VPG#BPMU (53% identity in 151 amino acids)

8065 SEQ ID NO: - : -0.821875, 65, a potential protease protein, similar to gpI [Bacteriophage Mu] gi|7226336|gb|aaF41489.1| (31% identity in 369 amino acids),

8070 SEQ ID NO: 1686 : -0.40171, 118, a putative major head subunit, similar to proteinT [Bacteriophage Mu] gi|6010409|gb|aaF01112.1|AF083977#31 (66% identity in 311 amino acids); hypothetical proteins for example ,[Neisseria meningitidis] gi|6900377|emb|CAB71989.1|(50% identity in 311 amino acids)

8075 SEQ ID NO: 1687 : -0.015888, 108, novel, similar to gp35 [Bacteriophage Mu] gi|6010410|gb|aaF01113.1|AF083977#32 (40% identity in 62 amino acids)

8080 SEQ ID NO: 1533 : -0.455151, 331, novel, similar to hypothetical proteins for example ,gp36 [Bacteriophage Mu] gi|6010411|gb|aaF01114.1|AF083977#33 (46% identity in 139 amino acids)

SEQ ID NO: 1534 : -0.761539, 105, novel, similar to hypothetical proteins for example ,gp37 [Bacteriophage Mu]

gi|1175870|sp|P44231|YF09#HAEIN (33% identity in 187 amino acids)

SEQ ID NO: 1535 : -0.293125, 161, novel, similar to hypothetical proteins for example ,gp38 [Bacteriophage Mu]

8085 gi|6010413|gb|aaF01116.1|AF083977#35 (54% identity in 52 amino acids)

SEQ ID NO: 1536: -0.370046, 218, a major tail subunit (sheath protein), similar to sheath protein GpL [Bacteriophage Mu]

gi|1834291|dbj|Baa19195.1| (51% identity in 499 amino acids); hypothetical proteins for example ,[Haemophilus influenzae Rd]

8090 gi|1175872|sp|P44233|YF11#HAEIN (40% identity in 499 amino acids)

SEQ ID NO: 1564 : -0.396053, 77, novel, similar to hypothetical proteins for example ,GpM [Bacteriophage Mu]

8095 gi|1834292|dbj|Baa19196.1| (49% identity in 120 amino acids)

SEQ ID NO: 1565 : -0.199849, 663, novel, similar to hypothetical proteins for example ,ORF3 [Bacteriophage Mu]

gi|1834293|dbj|Baa19197.1| (49% identity in 122 amino acids)

8100 [0027]

10) Proteins processing DNA/RNA

Sequence number: hydrophobicity, The number of amino acids, Character such as function

SEQ ID NO: 1395: -0.645885, 803, a type III secretion protein (surfacepresentation of antigens), similar to N-terminal part of putative type III secretion proteins for example ,SpaR protein (surface presentation of antigens) [Salmonella typhimurium]

8105 gi|730799|sp|P40701|SPAR#SALTY(44% identity in 144 amino acids), probably interrupted

8110 SEQ ID NO: 1396: -0.414798, 224, a type III secretion protein, similar to type IIIsecretion proteins for example ,SpaQ [Salmonella enterica]

gi|975756|gb|aaC43847.1| (68% identity in 86 amino acids)

SEQ ID NO: 1397: -0.230128, 157, type III secretion protein,

8115 similar to type III secretion proteins for example ,SpaP
[Salmonella enterica] gi|975755|gb|aaC43846.1| (69% identity
in 218 amino acids)
SEQ ID NO: 1398: 0.60339, 60, type III secretion protein,
similar to type III secretion proteins for example ,SpaO
8120 [Salmonella enterica] gi|973277|gb|aaC43944.1| (32% identity
in 292 amino acids)
SEQ ID NO: 1399: -0.623677, 795, type III secretion protein,
similar to C-terminal part of type III secretion proteins for
example ,SpaN [Salmonella enterica]
8125 gi|1155289|gb|aaC44993.1| (38% identity in 82 amino acids),
TTG start
SEQ ID NO: 1400: -0.241304, 47, novel
SEQ ID NO: - : -0.288136, 60, a type III secretion protein,
similar to type III secretion proteins for example ,SpaM
8130 [Salmonella enterica] gi|1155297|gb|aaC44998.1| (29%
identity in 146 amino acids)
SEQ ID NO: 1412: -0.074167, 361, a putative tape measure
protein, similar to hypothetical proteins for example ,Gp42
(putative tape measure protein) [Bacteriophage Mu]
8135 gi|6010417|gb|aaF01120.1|AF083977#39 (36% identity in 686
amino acids)
SEQ ID NO: 1413: -0.064607, 357, a putative DNA circulation
protein, similar to DNA circulation proteins for example ,
protein N [Bacteriophage Mu]
8140 gi|6010418|gb|aaF01121.1|AF083977#40 (33% identity in 441
amino acids)
SEQ ID NO: 1414: -0.374289, 845, a putative tail protein,
similar to tail protein for example ,P protein
[Bacteriophage Mu] gi|139353|sp|P08558|VPP#BPMU (47%
8145 identity in 360 amino acids), GTG start
SEQ ID NO: 1415: 0.2, 54, novel, similar to hypothetical
proteins for example ,gp45 [Bacteriophage Mu]
gi|6010420|gb|aaF01123.1|AF083977#42 (51% identity in 195

amino acids), may be involved in base plate assembly

8150 SEQ ID NO: 1416 : -0.05748, 128, novel, similar to hypothetical proteins for example ,Gp46 [Bacteriophage Mu] gi|6010421|gb|aaF01124.1|AF083977#43 (53% identity in 144 amino acids)

SEQ ID NO: 1417: -0.1945, 201, novel, similar to hypothetical

8155 proteins for example ,Gp47 [Bacteriophage Mu] gi|6010422|gb|aaF01125.1|AF083977#44 (53% identity in 360 amino acids)

SEQ ID NO: 1661: -0.169, 301, novel, similar to hypothetical proteins for example ,Gp48 [Bacteriophage Mu]

8160 gi|6010423|gb|aaF01126.1|AF083977#45 (48% identity in 183 amino acids)

SEQ ID NO: 1556 : -0.241844, 283, a putative tail fiber, similar to S protein [Bacteriophage Mu] gi|6010424|gb|aaF01127.1|AF083977#46 (46% identity in 198 amino acids); hypothetical proteins for example ,Bcv [Shigella boydii] gi|96900|pir|A42463 (56% identity in 78 amino acids)

8165 SEQ ID NO: 1557 : 0.691919, 100, a putative tail fiber assembly protein, similar to unnamed protein product [Bacteriophage 186] gi|3522882|gb|aaC34165.1| (39% identity in 173 amino acids); tail fiber assembly proteins for example ,U protein[Bacteriophage Mu] gi|6010425|gb|aaF01128.1|AF083977#47 (28% identity in 176 amino acids)

8175 SEQ ID NO: 1667: 1.052233, 292, similar to a C-terminal part of tail fiber protein (partial), C-terminal part of tail fiber proteins for example ,S [Bacteriophage Mu] gi|6010424|gb|aaF01127.1|AF083977#46 (38% identity in 65 amino acids)

8180 SEQ ID NO: - : -0.43064, 298, a putative invertase, similar to site-specific recombinases for example ,DNA-invertas for example ,in [Bacteriophage Mu]

gi|6010426|gb|aaF01129.1|AF083977#50 (75% identity in 181 amino acids)

8185 SEQ ID NO: 1600 -0.069079, 305, novel, similar to hypothetical proteins for example ,L0105 [Bacteriophage 933W] gi|4585419|gb|aaD25447.1|AF125520#42 (73% identity in 614 amino acids)

SEQ ID NO: - : -0.338889, 73, novel, similar to orf25 [Bacteriophage 933W] gi|4499806|emb|CAB39305.1| (52% identity in 57 amino acids)

8190 SEQ ID NO: 1616 : -0.524138, 465, novel, similar to hypothetical proteins for example ,L0106 [Bacteriophage 933W] gi|4585420|gb|aaD25448.1|AF125520#43 (41% identity in 79 amino acids)

8195 SEQ ID NO: 1630: -0.041597, 239, novel [0028]

11) Proteins relating pathogenicity

Sequence number: hydrophobicity, The number of amino acids.

8200 Character such as function

SEQ ID NO: 1631: 0.342857, 225, a type III secretion protein (ATP synthetase), similar to putative type III secretion proteins (ATP synthetase) for example ,invC [Salmonella typhimurium] gi|730791|sp|P39444|SPAL#SALTY (63% identity in 387 amino acids)

8205 SEQ ID NO: 1472: -0.763847, 1395, a type III secretion protein, similar to type III secretion proteins for example ,InvA [Salmonella typhimurium] gi|476819|pir||A42888 (64% identity in 686 amino acids)

8210 SEQ ID NO: - : -0.352577, 98, a type III secretion protein, similar to type III secretion proteins for example ,invasion protein [Salmonella enterica] gi|1236845|gb|aaC45041.1| (37% identity in 355 amino acids)

SEQ ID NO: 1552: -0.029639, 389, a type III secretion protein, similar to type III secretion proteins for example , InvG [Salmonella typhimurium]

8215

- gi|1170574|sp|P35672|INVG#SALTY (53% identity in 558 amino acids)
- SEQ ID NO: - , 0.760046, 439, a transcriptional regulator of type III secretion system, similar to transcriptional regulators for example ,invF [Salmonella typhimurium]
- 8220 gi|729852|sp|P39437|INVF#SALTY (40% identity in 200 amino acids)
- SEQ ID NO: 690: -0.029412, 52, novel, GTG start
- 8225 SEQ ID NO: 691: -0.113448, 410, novel, GTG start
- SEQ ID NO: 692 : 0.817973, 218, novel, similar to hypothetical proteins for example ,[Methanobacterium thermoautotrophicum] gi|7482365|pir||D69031 (32% identity in 100 amino acids)
- 8230 SEQ ID NO: 693 : -0.541477, 177, a putative transporter, similar to hypothetical protein [plasmid pNZ4000] gi|5123516|gb|aaD40355.1| (31% identity in 185 amino acids); similar to (at low level) putative low-affinity inorganic phosphate transporter [Mycobacterium tuberculosis]
- 8235 gi|7387993|sp|O06411|PIT#MYCTU (26% identity in 212 amino acids)
- SEQ ID NO: 694: -0.540244, 83, a hypothetical lipoprotein, similar to hypothetical proteins for example ,[plasmid pNZ4000] gi|5123517|gb|aaD40356.1|(25% identity in 209 amino acids)
- 8240 SEQ ID NO: 695: -0.645115, 697, a putative ABC transporter ATP-bindingsubunit, similar to ABC transporter ATP-binding subunits for example ,cation ABC transporter (ATP-binding protein) homolog ykoD [Bacillus subtilis]
- 8245 gi|7445788|pir||H69858 (32% identity in 201 amino acids)
- SEQ ID NO: 696: -0.096774, 311, a putative ABC-transporter ATP-bindingsubunit, similar to ABC-transporter ATP-binding subunits for example ,PotA homolog [Agrobacterium rhizogenes plasmid pRi1724] gi|8918682|dbj|Baa97747.1|
- 8250 (35% identity in 223 amino acids); [plasmid pNZ4000]

gi|5123519|gb|aaD40358.1| (30% identity in 211 amino acids)
 SEQ ID NO: 697 : 0.076712, 74, novel, similar to
 YGGC#ECOLI gi|1789296 (83% identity in 233 amino acids),
 but comprising different C-terminal part.

8255 : similar to kinaselike protein FrcK [Sinorhizobium meliloti]
 dad|AF196574-5|aaG28501.1 (38% identity in 190 amino acids),
 GTG start
 SEQ ID NO: 698 : -0.44881, 85, novel (hypothetical
 lipoprotein)

8260 SEQ ID NO: 699 : -0.246237, 94, a integrase, similar to
 integrases for example ,[prophage P4]
 gi|6179516|emb|CAB59974.1| (55% identity in 414 amino
 acids)
 SEQ ID NO: 700: -0.042222, 91, novel, similar to C-terminal
 part of hypothetical proteins for example ,L0015
 [Escherichia coli O-157:H7 strain EDL933]]
 gi|4808945|gb|aaD30027.1|AF119170#2(88% identity in 206
 amino acids), GTG start, probably disrupted

8265 SEQ ID NO: 701: -0.378351, 98, novel, similar to a part of
 hypothetical proteins for example ,L0013 [Escherichia coli
 O-157:H7 strain EDL933] gi|3414881|gb|aaC31492.1| (100%
 identity in 44 amino acids), GTG start, probably disrupted

8270 SEQ ID NO: 702 : -0.572727, 177, novel, similar to
 hypothetical proteins for example ,ORF29 [Escherichia coli
 plasmid pB171] gi|6009405|dbj|Baa84864.1| (39% identity in
 204 amino acids)

8275 SEQ ID NO: 703 : -0.159444, 181, novel, similar to
 hypothetical proteins for example ,ORF30 [Escherichia coli
 plasmid pB171] gi|6009406|dbj|Baa84865.1| (80% identity in
 115 amino acids)

8280 SEQ ID NO: 704 : 0.131638, 178, novel, similar to
 hypothetical proteins for example ,ORF31 [Escherichia coli
 plasmid pB171] gi|6009427|dbj|Baa84886.1| (63% identity in
 468 amino acids)

8285 SEQ ID NO: 705 : -0.321053, 172, novel, similar to
hypothetical protein [Salmonella choleraesuis]
gi|7467227|pir|T28668 (43% identity in 261 amino acids)
SEQ ID NO: 706 : -0.725, 97, a putative virulence-related
membrane protein, similar to virulence-related membrane
8290 proteins for example ,pagC [Salmonella typhimurium]
gi|129558|sp|P23988|PAGC#SALTY (45% identity in 171
amino acids)
SEQ ID NO: 707 : -0.170161, 125, novel
SEQ ID NO: 708 : -1.030769, 66, novel, similar to (at low level)
8295 hypothetical proteins for example ,FhaB [Neisseria
meningitidis] gi|6900333|emb|CAB71945.1| (37% identity in
48 amino acids), GTG start
SEQ ID NO: 709 : 0.1, 99, novel, identical to L0028
[Escherichia coli O-157:H7 strain EDL933]
8300 gi|3414896|gb|aaC31507.1|; similar to hypothetical proteins
for example ,[Escherichia coli] gi|3249026|gb|aaC69313.1|
(99% identity in 203 amino acids)
SEQ ID NO: 710 : -0.514201, 170, novel, identical to L0029
[Escherichia coli O-157:H7 strain EDL933]
8305 gi|3414897|gb|aaC31508.1|; similar to rOrf10 [Escherichia
coli] gi|2865295|gb|aaC38388.1| (78% identity in 119 amino
acids)
SEQ ID NO: 711 : -0.516312, 142, novel, identical to L0030
[Escherichia coli O-157:H7 strain EDL933]
8310 gi|3414898|gb|aaC31509.1|; similar to Orf18 [Escherichia
coli] gi|2865294|gb|aaC38387.1| (74% identity in 159 amino
acids)
SEQ ID NO: 712 : -0.221687, 167, a type III secretion system
SepQ protein, identical to L0031 [Escherichia coli O-157:H7
8315 strain EDL933]; gi|3414899|gb|aaC31510.1|; similar to SepQ
[Escherichia coli strain E2348/69] gi|2865293|gb|aaC38386.1|
(93% identity in 305 amino acids)
SEQ ID NO: 713 : -0.647059, 86, novel, similar to Orf16

- 8320 [Escherichia coli strain E2348/69] gi|2865292|gb|aaC38385.1| (97% identity in 138 amino acids); L0032 [Escherichia coli O-157:H7 strain EDL933] gi|3414900|gb|aaC31511.1| (100% identity in 91 amino acids)
SEQ ID NO: 714: -0.245946, 149, novel, identical to L0033 [Escherichia coli O-157:H7 strain EDL933]
- 8325 gi|3414901|gb|aaC31512.1|
SEQ ID NO: 715: -0.574667, 76, a type III secretion system protein EscN, identical to EscN (L00349 [Escherichia coli O-157:H7 strain EDL933] gi|3414902|gb|aaC31513.1|
SEQ ID NO: 716: -0.092157, 103, a type III secretion system
- 8330 EscV protein, identical to EscV (L0035) [Escherichia coli O-157:H7 strain EDL933] gi|3414903|gb|aaC31514.1|
SEQ ID NO: 717: -0.296875, 97, novel, identical to Orf12 [Escherichia coli strain E2348/69] gi|2865288|gb|aaC38381.1|;
L0036 [Escherichia coli O-157:H7 strain EDL933]
- 8335 gi|3414904|gb|aaC31515.1|
SEQ ID NO: 718: -0.570466, 194, identical to type III secretion system SepZ protein, SepZ proteins for example, [Escherichia coli O-157:H7 strain EDL933] gi|3414905|gb|aaC31516.1|
- 8340 SEQ ID NO: 719: -0.367148, 555, novel, identical to L0038 [Escherichia coli O-157:H7 strain EDL933] gi|3414906|gb|aaC31517.1|; similar to rOrf8 [Escherichia coli E2348/69] gi|2865287|gb|aaC38380.1| (92% identity in 142 amino acids)
- 8345 SEQ ID NO: 720: -0.356102, 509, a type III secretion system EscJ protein, identical to EscJ [Escherichia coli strain E2348/69] gi|2865286|gb|aaC38379.1|; L0039 (EscJ) [Escherichia coli O-157:H7 strain EDL933] gi|3414907|gb|aaC31518.1|
- 8350 SEQ ID NO: 721: -0.399319, 442, a type III secretion system protein SepD, identical to SepD (L0040) [Escherichia coli O-157:H7 strain EDL933] gi|3414908|gb|aaC31519.1|; similar

to SepD proteins for example, [Escherichia coli strain E2348/69] gi|886476|emb|Caa90273.1| (98% identity in 151 amino acids)

8355 SEQ ID NO: 722: -0.538854, 158, a type III secretion system EscC protein, identical to EscC (L0041) [Escherichia coli O-157:H7 strain EDL933] gi|3414909|gb|aaC31520.1|

8360 SEQ ID NO: 723: -0.272994, 375, a type III secretion system CesD protein, identical to CesD (L0042) [Escherichia coli O-157:H7 strain EDL933] gi|3414910|gb|aaC31521.1|

8365 SEQ ID NO: 724: -0.223492, 316, novel, identical to L0043 [Escherichia coli O-157:H7 strain EDL933] gi|3414911|gb|aaC31522.1|; similar to Orf11 [Escherichia coli strain E2348/69] gi|2865282|gb|aaC38375.1| (98% identity in 137 amino acids)

8370 SEQ ID NO: 725: -0.455469, 129, novel, identical to L0044 [Escherichia coli O-157:H7 strain EDL933] gi|3414912|gb|aaC31523.1|; similar to Orf10 [Escherichia coli strain E2348/69] gi|2865281|gb|aaC38374.1| (98% identity in 123 amino acids)

8375 SEQ ID NO: 726: -0.330216, 140, novel, identical to L0045 [Escherichia coli O-157:H7 strain EDL933] gi|3414913|gb|aaC31524.1|; similar to rOrf3 [Escherichia coli strain E2348/69] gi|2865280|gb|aaC38373.1| (98% identity in 152 amino acids)

8380 SEQ ID NO: 727: -0.154301, 187, a type III secretion system EscU protein, identical to EscU (L0046) [Escherichia coli O-157:H7 strain EDL933] gi|3414914|gb|aaC31525.1|

8385 SEQ ID NO: 728: -0.764198, 82, a type III secretion system EscT protein, identical to EscT (L0047) [Escherichia coli O-157:H7 strain EDL933] gi|3414915|gb|aaC31526.1|

SEQ ID NO: 729: -0.1374, 501, a type III secretion system EscS protein, identical to EscS (L0048) [Escherichia coli O-157:H7 strain EDL933] gi|3414916|gb|aaC31527.1|

SEQ ID NO: 730: -0.500827, 122, a type III secretion system

EscR protein, identical to EscR (L0049) [Escherichia coli O-157:H7 strain EDL933| gi|3414917|gb|aaC31528.1|

SEQ ID NO: 731: -0.213291, 159, novel, identical to L0050

8390 [Escherichia coli O-157:H7 strain EDL933| gi|3414918|gb|aaC31529.1|; similar to Orf5 [Escherichia coli strain E2348/69| gi|2865275|gb|aaC38368.1| (98% identity in 231 amino acids)

SEQ ID NO: 732: -0.205065, 692, novel, identical to L0051

8395 [Escherichia coli O-157:H7 strain EDL933| gi|3414919|gb|aaC31530.1|; similar to Orf4 [Escherichia coli strain E2348/69| gi|2865274|gb|aaC38367.1| (99% identity in 199 amino acids)

SEQ ID NO: 733: -0.131141, 457, novel, identical to Orf3

8400 [Escherichia coli E2348/69|gi|2865273|gb|aaC38366.1|; L0052 [Escherichia coli O-157:H7 strain EDL933| gi|3414920|gb|aaC31531.1|

SEQ ID NO: 734: -0.277807, 375, novel, similar to Orf2 [Escherichia coli strain E2348/69| gi|2865272|gb|aaC38365.1|

8405 (98% identity in 72 amino acids); L0053 [Escherichia coli O-157:H7 strain EDL933| gi|3414921|gb|aaC31532.1| (98% identity in 72 amino acids)

SEQ ID NO: 735: -0.335784, 205, a transcription regulatory element, identical to L0054 [Escherichia coli O-157:H7 strain EDL933| gi|3414922|gb|aaC31533.1| ; similar to Orf1 Ler [Escherichia coli strain E2348/69| gi|2865271|gb|aaC38364.1| (99% identity in 129 amino acids)

SEQ ID NO: 736: -0.142069, 146, novel

SEQ ID NO: 737: -0.199169, 362, a secreted effector protein,

8415 identical to L0055 [Escherichia coli O-157:H7 strain EDL933| gi|3414923|gb|aaC31534.1|; similar to rOrf2 EspG [Escherichia coli strain E2348/69| gi|2865270|gb|aaC38363.1| (97% identity in 398 amino acids)

SEQ ID NO: 738: -0.374731, 187, novel, identical to L0056

8420 [Escherichia coli O-157:H7 strain EDL933|

gi|3414924|gb|aaC31535.1| : similar to rOrf1 [Escherichia
colistrain E2348/69] gi|2865269|gb|aaC38362.1| (99% identity
in 272 amino acids)

SEQ ID NO: 739: -0.368977, 304, novel, TTG start

8425 SEQ ID NO: 740: -0.53815, 174, novel

SEQ ID NO: 741: -0.097015, 68, novel, similar to hypothetical
proteins for example ,NMA0565 [Neisseria meningitidis]
gi|7379302|emb|CAB83857.1 (35% identity in 118 amino acids)

SEQ ID NO: 742: -0.458602, 187, novel

8430 SEQ ID NO: 743: -0.278645, 680, a putative transcriptional[sic,
translational]regulator , similar to transcriptional[sic,
translational] regulators for example ,Com protein
(transcriptional[sic, translational] regulator of Mom)
[Bacteriophage Mu] gi|7388376|sp|Q53979|VCOM#SHIDY(46%
8435 identity in 57 amino acids)

SEQ ID NO: 744: 0.096667, 61, a putative DNA modification
protein, similar to DNA modification proteins for
example ,Mom protein [Bacteriophage Mu]
gi|138782|sp|P06018|VMOM#BPMU (76% identity in 245
8440 amino acids), GTG start

SEQ ID NO: 745: -0.679012, 82, a sorbose-permease IID
component(PTS system), similar to sorbose-permease IID
components for example ,[Klebsiella pneumoniae]
gi|548634|sp|P37083|PTRD#KLEPN (92% identity in 64 amino
8445 acids), interrupted byphage-insertion

SEQ ID NO: 746: -0.529126, 104, a sorbose-permease IIC
component (PTS system), similar to sorbose-permease IIC
components for example ,[Klebsiella pneumoniae]
gi|548633|sp|P37082|PTRC#KLEPN (92% identity in 265
8450 amino acids)

SEQ ID NO: 747: -0.937879, 67, a sorbose-permease IIB
component (PTS system), similar to sorbose-permease IIB
components for example ,[Klebsiella pneumoniae]
gi|1142714|gb|aaB04152.1| (46% identity in 162 amino acids)

- 8455 SEQ ID NO: 748: -0.563673, 246, a putative sorbose-permease IIA component (PTS system), similar to sorbose-permease IIA components, for example ,[Klebsiella pneumoniae] gi|548631|sp|P37080|PTR#KLEPN (71% identity in 135 amino acids)
- 8460 SEQ ID NO: 749: -0.055385, 66, a sorbitol-6-phosphate 2-dehydrogenase, similar to sorbitol-6-phosphate 2-dehydrogenases for example ,[Klebsiella pneumoniae] gi|548951|sp|P37079|SORD#KLEPN (86% identity in 268 amino acids)
- 8465 SEQ ID NO: 750: 0.997359, 266, a putative sorbitol operon regulatory element (activator), similar to sorbitol operon regulatory element (SorC family) for example ,[Klebsiella pneumoniae] gi|548950|sp|P37078|SORC#KLEPN (86% identity in 315 amino acids)
- 8470 SEQ ID NO: 751: -0.115244, 165, a putative regulatory protein, similar to regulatory proteins for example ,aerobic respiration control protein [Zymomonas mobilis] gi|4511977|gb|aaD21537.1| (39% identity in 230 amino acids)
- 8475 SEQ ID NO: 752: 0.19037, 136, a putative sugar kinase, similar to sugar kinases for example ,fructo kinase homolog ydJ [Bacillus subtilis] gi|3915420|sp|O34768|YDJE#BACSU (24% identity in 326 amino acids)
- SEQ ID NO: 753: -0.159702, 269, a putative aldolase, similar to aldolases for example ,fructose-bisphosphate aldolase (EC 4.1.2.13) Fbaa [Bacillus subtilis] gi|543796|sp|P13243|ALF1#BACSU (41% identity in 286 amino acids)
- 8480 SEQ ID NO: 754: -0.218413, 316, novel, similar to (at low level) a part of hypothetical protein ydaE [Bacillus subtilis] gi|7474928|pir|E69768 (35% identity in 51 amino acids)
- 8485 SEQ ID NO: 1322: 0.197872, 236, a putative carbohydratebinding protein, similar to C-terminal part of carbohydratebinding proteins for example , bifunctional

- carbohydrate binding and transporter protein [Streptomyces
 8490 coelicolor A3(2)] gi|6714794|emb|CAB66286.1| (35% identity
 in 304 amino acids); ribose ABC transporter (ribose-binding
 protein) rbsB [Bacillus subtilis]
 gi|6174949|sp|P36949|RBSB#BACSU(36% identity in 261
 amino acids)
- 8495 SEQ ID NO: 1323: -0.163964, 334, a putative carbohydrate
 ABC transporter (permease), similar to carbohydrate ABC
 transporters (permease) for example ,ribose ABC transporter
 (permease) rbsC [Bacillus subtilis] gi|7446897|pir|B69690
 (43% identity in 317 amino acids)
- 8500 SEQ ID NO: 1324: 0.066434, 287, a putative sugar ABC
 transporter, ATP-binding protein, similar to sugar ABC
 transporter, ATP-binding proteins for example ,riboseABC
 transporter (ATP-binding protein) rbsA [Bacillus subtilis]
 gi|7404442|sp|P36947|RBSA#BACSU (45% identity in 489
 8505 amino acids)
- SEQ ID NO: 1325: -0.440969, 228, a putative histidine
 protein kinase, similar to histidine protein kinase for
 example ,histidine protein kinase-response regulator hybrid
 protein CvgSY [Pseudomonas syringae pv. syringae]
 8510 gi|5019771|gb|aaD37857.1|AF133263#2 (43% identity in 364
 amino acids)
- SEQ ID NO: 1326: -0.003195, 314, a putative transposase,
 similar to transposase homologA [Helicobacter pylori]
 gi|2114470|gb|aaD11513.1 (60% identity in 137 amino acids)
- 8515 SEQ ID NO: 1327: 1.026235, 325, a putative transposase,
 similar to B1432#ECOLI gi|1787702 (96% identity in 402 amino
 acids); transposases for example ,ORFB [Xylella fastidiosa]
 gi|9105393|gb|aaF83346.1|AE003901#9 (38% identity in 321
 amino acids)
- 8520 SEQ ID NO: 1328: -0.04664, 507, a putative integrase,
 similar to(at low level) integrases for example ,integrase
 [Bacteriophage TPW22]

- gi|6465906|gb|aaF12706.1|AF066865#4 (23% identity in 342 amino acids)
- 8525 SEQ ID NO: - : -0.010053, 757, identical to transposase (insertion sequence IS629),gi|7444868|pir||T00241
SEQ ID NO: 1620: -0.25035, 144, identical to transposase (insertion sequence IS629),[Escherichia coli plasmid p O-157] gi|7443862|pir||T00240
- 8530 SEQ ID NO: 1621: -0.587696, 383, novel
SEQ ID NO: 1310: -0.455932, 650, novel, TTG start
SEQ ID NO: 1311: -0.965741, 109, novel, TTG start
SEQ ID NO: 1312: -0.397973, 297, novel, similar to(at low level) hypothetical proteins [Staphylococcus aureus] for example ,gi|7594765|dbj|Baa94663.1| (30% identity in 143 amino acids); hypothetical protein [Neisseria meningitidis] gi|5051461|emb|CAB44981.1| (28% identity in 140 amino acids)
SEQ ID NO: 1313: -0.511702, 95, a putative resolvase, similar to resolvases for example ,resolvase [Escherichia coli transposon Tn2501] gi|135944|sp|P05823|TNP0#ECOLI(45% identity in 179 amino acids)
[0029]
- 12) Other proteins
- 8545 Sequence number: hydrophobicity. The number of amino acids. Character such as function
SEQ ID NO: 1314:0.037273, 111, putative transposase, similar to C-terminal part of transposases, for example, [Escherichia coli Tn5] gi|622948|gb|aaB60064.1|, may be disrupted
- 8550 SEQ ID NO: 1315: -0.213793, 59, novel, similar to a part of KfaE protein [Escherichia coli] gi|628752|pir||S45104 (55% identity in 52 amino acids)
SEQ ID NO: 1316: -0.256129, 156, a putative enterotoxin, similar to ShET2 enterotoxin [Shigella flexneri] gi|1109754|emb|Caa90938.1| (38% identity in 539 amino acids) ; similar to a part of hypothetical protein, for example,

ankyri-like regulatory protein [Escherichia coli]
gi|418526|sp|P23325|ARP#ECOLI (28% identity in 172 amino acids) (at low level)

8560 SEQ ID NO: 1317: -0.050262, 192, novel, similar to sB protein, for example, [insertion element iso-IS1N] gi|124919|sp|P03832|ISBN#SHIDY (69% identity in 49 amino acids), TTG start

SEQ ID NO: 1318: -0.438356, 439, novel, similar to a

8565 hypothetical protein [Salmonella typhimurium] gi|6960367|gb|aaF33527.1| (63% identity in 306 amino acids)

SEQ ID NO: 1319: -0.524125, 258, novel

SEQ ID NO: 1320: -0.435714, 155, novel, similar to a hypothetical protein in insertion elements, for example, [IS630]

8570 gi|140943|sp|P16943|YIS5#SHISO (88% identity in 282 amino acids)

SEQ ID NO: 1014: -0.510181, 276, a putative adherence factor, similar to N-terminal part of adherence factors (amino acids at the position 1-433/3223), for example, Efa1 [Escherichia coli

8575 O111:H- strain E45035] gi|6013469|gb|aaD49229.2|AF159462#1 (99% identity in 433 amino acids), probably interrupted by frameshift

SEQ ID NO: 1015: -0.496819, 284, a putative DNA-binding protein, similar to putative DNA-binding protein [Neisseria meningitidis]

8580 gi|7379301|emb|CAB83856.1 (47% identity in 101 amino acids)

SEQ ID NO: 1016: -0.412037, 109, novel

SEQ ID NO: 1017: -0.505722, 368, a putative transcription regulatory element, its N-terminal part is similar to

8585 transcription regulatory elements, for example, BamH I control element [Bacillus amyloliquefaciens] gi|116073|sp|P23939|CEBA#BACAM (47% identity in 68 amino acids)

SEQ ID NO: 1018: -0.409362, 236, an integrase, similar to

8590 integrase, for example, [prophage P4]

gi|732036|sp|P39347|INTB#ECOLI (74% identity in 236 amino acids)
 SEQ ID NO: 1019: -0.205818, 551, novel
 SEQ ID NO: 1020: -0.198657, 1118, novel, similar to a part of
 8595 hypothetical proteins, for example, YjH [Escherichia coli|
 gi|7404491|sp|P39371|YJHT#ECOLI (95% identity in 82
 amino acids), TTG start
 SEQ ID NO: 1021: -0.398339, 2105, novel
 SEQ ID NO: 1022: -0.508378, 944, novel, similar to putative
 8600 periplasmic protein [Campylobacter jejuni|
 gi|6968066|emb|CAB75235.1| (26% identity in 173 amino
 acids) (at low level)
 SEQ ID NO: 1023: -0.482301, 1645, novel (putative membrane
 protein), similar to a part of myosin heavy chains, for example,
 8605 [Cyprinus carpio] gi|2351223|dbj|Baa22069.1| (19% identity
 in 292 amino acids) (at low level)
 SEQ ID NO: 1024: -0.359727, 2114, novel, similar to a part of
 YjiT [Escherichia coli] gi|732099|sp|P39391|YJIT#ECOLI
 (27% identity in 239 amino acids) (at low level), GTG start
 8610 SEQ ID NO: 1025: -0.345738, 705, novel, its N-terminal part is
 similar to N-terminal part of putative RNA helicase
 [Deinococcus radiodurans (strain R1)] gi|7473663|pir|B75633
 (29% identity in 291 amino acids); and its central part is
 similar to hypothetical YjiV protein [Escherichia coli]
 8615 gi|2851665|sp|P39393|YJIV#ECOLI (28% identity in 491
 amino acids); a part of McrD protein [Escherichia coli]
 gi|2851619|sp|P27301|MCRD#ECOLI (39% identity in 131
 amino acids)
 SEQ ID NO: 1026: 0.04, 61, a putative ATP-dependent helicase,
 8620 similar to putative ATP-dependent helicases, for example,
 [Halobacterium sp. (strain NRC-1) plasmid pNRC100]
 gi|7484100|pir|T08316 (26% identity in 597 amino acids)
 SEQ ID NO: 1027: -0.514474, 77, novel, similar to hypothetical
 proteins, for example, H1130 [Halobacterium sp. (strain NRC-1)]

- 8625 plasmid pNRC100| gi|7484076|pir||T08313 (25% identity in 508 amino acids); and possible restriction /modificationenzyme [Campylobacter jejuni| gi|6968147|emb|CAB72964.1| (24% identity in 414 amino acids)
SEQ ID NO: 1028: -0.40375, 81, a putative RNA helicase,
- 8630 similar to putative RNA helicases, for example, [Deinococcus radiodurans (strain R1)| gi|7473663|pir||B75633 (amino acids at the position 78-396) (31% identity in 318 amino acids); and (amino acids at the position 994-1708) (23% identity in 714 amino acids)
- 8635 SEQ ID NO: 1468: -0.351742, 1580, a putative DNA helicase, similar to DNA helicases, for example, putative DNA helicase H91#ORF529 [Mycoplasma pneumoniae| gi|2495150|sp|P75438|YH91#MYCPN (24% identity in 455 amino acids); and helicase IV [Escherichia coli|
- 8640 gi|146328|gb|aaA23952.1| (23% identity in 513 amino acids)
SEQ ID NO: 1469: 0.14127, 64, novel, TTG start
SEQ ID NO: 1470: -0.245455, 67, novel, similar to N-terminal part of putative membrane protein b1978 [Escherichia coli K-12| gi|1736642|dbj|Baa15799.1| (58% identity in 46 amino
- 8645 acids)
SEQ ID NO: 1546: -0.622994, 736, novel
SEQ ID NO: - : -0.059091, 89, novel
SEQ ID NO: 1592: -0.298976, 294, novel, similar to N-terminal part of hypothetical proteins, for example, jhp0462
- 8650 [Helicobacter pylori (strain J99)| gi|7464730|pir||C71929 (48% identity in 269 amino acids); and jhp0572 [Helicobacter pylori (strainJ99)| gi|7464757|pir||H71914 (31% identity in 282 amino acids)
SEQ ID NO: 1593: -0.494832, 388, novel, similar to C-terminal
- 8655 part of hypothetical proteins, for example, jhp0462 [Helicobacter pylori (strain J99)| gi|7464730|pir||C71929 (42% identity in 423 amino acids); and HP051 3 [Helicobacter pylori (strain26695)| gi|7464291|pir||A64584 (44% identity in

423 amino acids)

- 8660 SEQ ID NO: 1381 : -0.367123, 585, a type I restriction
modification enzymeS subunit, similar to type I
restriction-modification enzyme S subunits, for example,
[Citrobacter freundii|pir|S06097| (54% identity in 584 amino
acids)
- 8665 SEQ ID NO: 1382 : -0.413184, 494, a type I restriction
modification enzymeM subunit, similar to type I restriction
modification enzyme M subunits, for example, [EcoA system]
gi|421016|pir||A47200 (98% identity in 489 amino acids)
- 8670 SEQ ID NO: 1383 : -0.505062, 811, a type I
restriction-modification enzymeR subunit, similar to type I
restriction-modification enzyme R subunits, for example, [EcoA]
gi|2121113|pir||I41291 (99% identity in 810 amino acids)
- 8675 SEQ ID NO: 1384: -0.614894, 95, novel, similar to N-terminal
part of hypothetical proteins, for example, [Helicobacter pylori]
gi|7464531|pir||E64694 (36% identity in 87 amino acids)
- SEQ ID NO: 1385 : -0.442477, 453, novel, similar to
hypothetical proteins, for example, [Streptomyces coelicolor
A3(2)] gi|7479715|pir||T35601 (22% identity in 379 amino
acids) (at low level), TTG start
- 8680 SEQ ID NO: 1689: -0.487222, 181, novel
[0030]

1) Proteins having unknown function

- These proteins or polypeptides are selected from a group
comprising the following sequence list: SEQ ID NO: 163, SEQ
- 8685 ID NO: 164, SEQ ID NO: 165, SEQ ID NO: 166, SEQ ID NO: 167,
SEQ ID NO: 168, SEQ ID NO: 169, SEQ ID NO: 170, SEQ ID
NO: 171, SEQ ID NO: 172, SEQ ID NO: 173, SEQ ID NO: 174,
SEQ ID NO: 175, SEQ ID NO: 176, SEQ ID NO: 177, SEQ ID
NO: 178, SEQ ID NO: 179, SEQ ID NO: 180, SEQ ID NO: 181,
- 8690 SEQ ID NO: 182, SEQ ID NO: 183, SEQ ID NO: 184, SEQ ID
NO: 185, SEQ ID NO: 186, SEQ ID NO: 187, SEQ ID NO: 188,
SEQ ID NO: 189, SEQ ID NO: 190, SEQ ID NO: 191, SEQ ID

NO: 192, SEQ ID NO: 193, SEQ ID NO: 194, SEQ ID NO: 195,
 SEQ ID NO: 196, SEQ ID NO: 197, SEQ ID NO: 198, SEQ ID
 8695 NO: 199, SEQ ID NO: 200, SEQ ID NO: 201, SEQ ID NO: 202,
 SEQ ID NO: 203, SEQ ID NO: 204, SEQ ID NO: 205, SEQ ID
 NO: 206, SEQ ID NO: 207, SEQ ID NO: 208, SEQ ID NO: 209,
 SEQ ID NO: 210, SEQ ID NO: 211, SEQ ID NO: 212, SEQ ID
 NO: 213, SEQ ID NO: 214, SEQ ID NO: 215, SEQ ID NO: 216,
 8700 SEQ ID NO: 217, SEQ ID NO: 218, SEQ ID NO: 219, SEQ ID
 NO: 220, SEQ ID NO: 221, SEQ ID NO: 222, SEQ ID NO: 223,
 SEQ ID NO: 224, SEQ ID NO: 225, SEQ ID NO: 226, SEQ ID
 NO: 227, SEQ ID NO: 228, SEQ ID NO: 229, SEQ ID NO: 230,
 SEQ ID NO: 231, SEQ ID NO: 232, SEQ ID NO: 233, SEQ ID
 8705 NO: 234, SEQ ID NO: 235, SEQ ID NO: 236, SEQ ID NO: 237,
 SEQ ID NO: 238, SEQ ID NO: 239, SEQ ID NO: 240, SEQ ID
 NO: 241, SEQ ID NO: 242, SEQ ID NO: 243, SEQ ID NO: 485,
 SEQ ID NO: 486, SEQ ID NO: 487, SEQ ID NO: 488, SEQ ID
 NO: 489, SEQ ID NO: 490, SEQ ID NO: 491, SEQ ID NO: 492,
 8710 SEQ ID NO: 493, SEQ ID NO: 494, SEQ ID NO: 495, SEQ ID
 NO: 496, SEQ ID NO: 497, SEQ ID NO: 498, SEQ ID NO: 499,
 SEQ ID NO: 500, SEQ ID NO: 501, SEQ ID NO: 502, SEQ ID
 NO: 503, SEQ ID NO: 504, SEQ ID NO: 505, SEQ ID NO: 506,
 SEQ ID NO: 507, SEQ ID NO: 508, SEQ ID NO: 509, SEQ ID
 8715 NO: 510, SEQ ID NO: 511, SEQ ID NO: 512, SEQ ID NO: 513,
 SEQ ID NO: 514, SEQ ID NO: 515, SEQ ID NO: 516, SEQ ID
 NO: 517, SEQ ID NO: 518, SEQ ID NO: 519, SEQ ID NO: 520,
 SEQ ID NO: 521, SEQ ID NO: 522, SEQ ID NO: 523, SEQ ID
 NO: 524, SEQ ID NO: 525, SEQ ID NO: 526, SEQ ID NO: 527,
 8720 SEQ ID NO: 528, SEQ ID NO: 529, SEQ ID NO: 530, SEQ ID
 NO: 531, SEQ ID NO: 532, SEQ ID NO: 533, SEQ ID NO: 534,
 SEQ ID NO: 535, SEQ ID NO: 536, SEQ ID NO: 537, SEQ ID
 NO: 538, SEQ ID NO: 539, SEQ ID NO: 540, SEQ ID NO: 541,
 SEQ ID NO: 542, SEQ ID NO: 543, SEQ ID NO: 544, SEQ ID
 8725 NO: 545, SEQ ID NO: 546, SEQ ID NO: 547, SEQ ID NO: 548,
 SEQ ID NO: 549, SEQ ID NO: 550, SEQ ID NO: 551, SEQ ID

NO: 552, SEQ ID NO: 553, SEQ ID NO: 631, SEQ ID NO: 632,
SEQ ID NO: 633, SEQ ID NO: 634, SEQ ID NO: 635, SEQ ID
NO: 636, SEQ ID NO: 637, SEQ ID NO: 638, SEQ ID NO: 639,
8730 SEQ ID NO: 640, SEQ ID NO: 641, SEQ ID NO: 642, SEQ ID
NO: 643, SEQ ID NO: 644, SEQ ID NO: 928, SEQ ID NO: 929,
SEQ ID NO: 930, SEQ ID NO: 931, SEQ ID NO: 932, SEQ ID
NO: 933, SEQ ID NO: 934, SEQ ID NO: 935, SEQ ID NO: 936,
SEQ ID NO: 937, SEQ ID NO: 938, SEQ ID NO: 939, SEQ ID
8735 NO: 940, SEQ ID NO: 941, SEQ ID NO: 942, SEQ ID NO: 943,
SEQ ID NO: 944, SEQ ID NO: 945, SEQ ID NO: 946, SEQ ID
NO: 979, SEQ ID NO: 980, SEQ ID NO: 981, SEQ ID NO: 982,
SEQ ID NO: 983, SEQ ID NO: 984, SEQ ID NO: 985, SEQ ID
NO: 986, SEQ ID NO: 987, SEQ ID NO: 988, SEQ ID NO: 989,
8740 SEQ ID NO: 990, SEQ ID NO: 991, SEQ ID NO: 992, SEQ ID
NO: 993, SEQ ID NO: 994, SEQ ID NO: 995, SEQ ID NO: 996,
SEQ ID NO: 997, SEQ ID NO: 998, SEQ ID NO: 999, SEQ ID
NO: 1000, SEQ ID NO: 1001, SEQ ID NO: 1002, SEQ ID NO:
1003, SEQ ID NO: 1004, SEQ ID NO: 1005, SEQ ID NO: 1006,
8745 SEQ ID NO: 1008, SEQ ID NO: 1009, SEQ ID NO: 1010, SEQ ID
NO: 1011, SEQ ID NO: 1012, SEQ ID NO: 1056, SEQ ID NO:
1057, SEQ ID NO: 1058, SEQ ID NO: 1059, SEQ ID NO: 1094,
SEQ ID NO: 1095, SEQ ID NO: 1096, SEQ ID NO: 1097, SEQ ID
NO: 1098, SEQ ID NO: 1099, SEQ ID NO: 1100, SEQ ID NO:
8750 1101, SEQ ID NO: 1102, SEQ ID NO: 1103, SEQ ID NO: 1104,
SEQ ID NO: 1105, SEQ ID NO: 1106, SEQ ID NO: 1107, SEQ ID
NO: 1108, SEQ ID NO: 1109, SEQ ID NO: 1110, SEQ ID NO:
1111, SEQ ID NO: 1112, SEQ ID NO: 1113, SEQ ID NO: 1114,
SEQ ID NO: 1115, SEQ ID NO: 1116, SEQ ID NO: 1117, SEQ ID
8755 NO: 1118, SEQ ID NO: 1119, SEQ ID NO: 1120, SEQ ID NO:
1121, SEQ ID NO: 1122, SEQ ID NO: 1123, SEQ ID NO: 1124,
SEQ ID NO: 1125, SEQ ID NO: 1126, SEQ ID NO: 1127, SEQ ID
NO: 1213, SEQ ID NO: 1214, SEQ ID NO: 1215, SEQ ID NO:
1216, SEQ ID NO: 1217, SEQ ID NO: 1218, SEQ ID NO: 1219,
8760 SEQ ID NO: 1220, SEQ ID NO: 1221, SEQ ID NO: 1222, SEQ ID

NO: 1223, SEQ ID NO: 1224, SEQID NO: 1225, SEQ ID NO:
 1226, SEQ ID NO: 1227, SEQ ID NO: 1228, SEQ IDNO: 1229,
 SEQ ID NO: 1230, SEQ ID NO: 1231, SEQ ID NO: 1232, SEQ ID
 NO:1233, SEQ ID NO: 1234, SEQ ID NO: 1235, SEQ ID NO:
 8765 1236, SEQ ID NO: 1237, SEQ ID NO: 1238, SEQ ID NO: 1239,
 SEQ ID NO: 1275, SEQ ID NO: 1276, SEQ ID NO: 1277, SEQ ID
 NO: 1278, SEQ ID NO: 1279, SEQ ID NO: 1280, SEQ IDNO:
 1281, SEQ ID NO: 1282, SEQ ID NO: 1284, SEQ ID NO: 1285,
 SEQ ID NO:1286, SEQ ID NO: 1287, SEQ ID NO: 1303, SEQ ID
 8770 NO: 1304, SEQ ID NO: 1305, SEQ ID NO: 1306, SEQ ID NO:
 1307, SEQ ID NO: 1308, SEQ ID NO: 1360, SEQ ID NO: 1361,
 SEQ ID NO: 1362, SEQ ID NO: 1363, SEQ ID NO: 1364, SEQ ID
 NO: 1365, SEQ ID NO: 1387, SEQ ID NO: 1388, SEQ ID NO:
 1389, SEQ ID NO: 1390, SEQ ID NO: 1391, SEQ ID NO: 1392,
 8775 SEQ ID NO: 1393, SEQ ID NO: 1437, SEQ ID NO: 1438, SEQ ID
 NO: 1439, SEQ ID NO: 1440, SEQ ID NO: 1441, SEQ ID NO:
 1442, SEQ ID NO: 1451, SEQ ID NO: 1452, SEQ ID NO: 1453,
 SEQ ID NO: 1454, SEQ ID NO: 1455, SEQ ID NO: 1456, SEQ ID
 NO: 1474, SEQ ID NO: 1475, SEQ ID NO: 1476, SEQ ID NO:
 8780 1479, SEQ ID NO: 1480, SEQ ID NO: 1481, SEQ ID NO: 1482,
 SEQ ID NO: 1483, SEQ ID NO: 1484, SEQ ID NO: 1485, SEQ ID
 NO: 1486, SEQ ID NO: 1495, SEQ ID NO: 1496, SEQ ID NO:
 1497, SEQID NO: 1498, SEQ ID NO: 1500, SEQ ID NO: 1502,
 SEQ ID NO: 1503, SEQ IDNO: 1504, SEQ ID NO: 1505, SEQ ID
 8785 NO: 1559, SEQ ID NO: 1560, SEQ ID NO:1561, SEQ ID NO:
 1562, SEQ ID NO: 1577, SEQ ID NO: 1578, SEQ ID NO: 1579,
 SEQ ID NO: 1602, SEQ ID NO: 1606, SEQ ID NO: 1625, SEQ ID
 NO: 1663, SEQ ID NO: 1697, SEQ ID NO: 1698, SEQ ID NO:
 1702 and SEQ ID NO: 1703. These proteins or polypeptides
 8790 are psecific to O-157:H7. Whereas no significant homology to
 all data registered in gene data bank is found from information
 of determined amino acid sequence, and their functions and the
 like are not known. However, as shown in table 1, a protein
 predicted to be a cell surface protein (membrane protein,

8795 especially, outer membrane protein (OMP), lipoprotein) in them
or its gene (or nucleic-acid molecule) may be useful for
production of an antibody, vaccine composition, diagnosis of
O-157 infection and the like. Furthermore, there is a
possibility that they include a protein which has an important
8800 function in O-157, for example, transportation and metabolism
of a substance, processing of nucleic acids, and relates to a
regulatory element and pathogenicity. They are to be useful
for diagnosis and therapy of O-157 infection.

[0031]

8805 2) Proteins which have unknown function, but have significant
homology to that of other bacteria:

These proteins or polypeptides are selected from a group
comprising the following sequence list: SEQ ID NO: 02, SEQ ID
NO: 03, SEQ ID NO: 04, SEQ ID NO: 05, SEQ ID NO: 06, SEQ ID
8810 NO: 07, SEQ ID NO: 08, SEQ ID NO: 09, SEQ ID NO: 10, SEQ
ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14,
SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18,
SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22,
SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26,
8815 SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30,
SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 245, SEQ ID NO:
246, SEQ ID NO: 247, SEQ ID NO: 248, SEQ ID NO: 249, SEQ
ID NO: 250, SEQ ID NO: 251, SEQ ID NO: 252, SEQ ID NO: 253,
SEQ ID NO: 254, SEQ ID NO: 255, SEQ ID NO: 256, SEQ ID
8820 NO: 257, SEQ ID NO: 258, SEQ ID NO: 259, SEQ ID NO: 260,
SEQ ID NO: 261, SEQ ID NO: 262, SEQ ID NO: 263, SEQ ID
NO: 264, SEQ ID NO: 265, SEQ ID NO: 266, SEQ ID NO: 267,
SEQ ID NO: 268, SEQ ID NO: 270, SEQ ID NO: 271, SEQ ID NO:
272, SEQ ID NO: 273, SEQ ID NO: 338, SEQ ID NO: 339, SEQ
8825 ID NO: 340, SEQ ID NO: 341, SEQ ID NO: 342, SEQ ID NO:
343, SEQ ID NO: 344, SEQ ID NO: 345, SEQ ID NO: 346, SEQ
ID NO: 347, SEQ ID NO: 348, SEQ ID NO: 349, SEQ ID NO: 350,
SEQ ID NO: 351, SEQ ID NO: 352, SEQ ID NO: 353, SEQ ID NO:

354, SEQ ID NO: 355, SEQ ID NO: 356, SEQ IDNO: 357, SEQ
 8830 ID NO: 358, SEQ ID NO: 359, SEQ ID NO: 360, SEQ ID NO:
 361,SEQ ID NO: 362, SEQ ID NO: 363, SEQ ID NO: 364, SEQ
 ID NO: 365, SEQ ID NO: 366, SEQ ID NO: 367, SEQ ID NO: 368,
 SEQ ID NO: 369, SEQ ID NO: 370,SEQ ID NO: 371, SEQ ID NO:
 372, SEQ ID NO: 373, SEQ ID NO: 374, SEQ ID NO: 375, SEQ
 8835 ID NO: 376, SEQ ID NO: 377, SEQ ID NO: 378, SEQ ID NO:
 379,SEQ ID NO: 380, SEQ ID NO: 381, SEQ ID NO: 382, SEQ
 ID NO: 383, SEQ ID NO: 384, SEQ ID NO: 385, SEQ ID NO: 386,
 SEQ ID NO: 387, SEQ ID NO: 388,SEQ ID NO: 389, SEQ ID NO:
 390, SEQ ID NO: 391, SEQ ID NO: 392, SEQ IDNO: 393, SEQ
 8840 ID NO: 394, SEQ ID NO: 395, SEQ ID NO: 396, SEQ ID NO: 397,
 SEQ ID NO: 398, SEQ ID NO: 399, SEQ ID NO: 400, SEQ ID
 NO: 401, SEQ IDNO: 402, SEQ ID NO: 403, SEQ ID NO: 404,
 SEQ ID NO: 405, SEQ ID NO: 406, SEQ ID NO: 407, SEQ ID
 NO: 408, SEQ ID NO: 409, SEQ ID NO: 411, SEQ ID NO: 412,
 8845 SEQ ID NO: 413, SEQ ID NO: 414, SEQ ID NO: 416, SEQ ID
 NO: 417, SEQ ID NO: 418, SEQ ID NO: 419, SEQ ID NO: 420,
 SEQ ID NO: 421, SEQ ID NO: 422, SEQ ID NO: 423, SEQ ID
 NO: 424, SEQ ID NO: 425, SEQ ID NO: 426, SEQ ID NO: 427,
 SEQ ID NO: 428, SEQ ID NO: 429, SEQ ID NO: 430, SEQ ID
 8850 NO: 431, SEQ ID NO: 432, SEQ ID NO: 433, SEQ ID NO: 434,
 SEQ ID NO: 435, SEQ ID NO: 436, SEQ ID NO: 437, SEQ ID
 NO: 438, SEQ ID NO: 439, SEQ ID NO: 440, SEQ ID NO: 441,
 SEQ ID NO: 442, SEQ ID NO: 443, SEQ ID NO: 444, SEQ ID
 NO: 445, SEQ ID NO: 446, SEQ ID NO: 447, SEQ ID NO: 448,
 8855 SEQ ID NO: 449, SEQ ID NO: 450, SEQ ID NO: 451, SEQ ID
 NO: 452, SEQ ID NO: 453, SEQ ID NO: 454, SEQ ID NO: 455,
 SEQ ID NO: 456, SEQ ID NO: 457, SEQ ID NO: 458, SEQ ID
 NO: 459, SEQ ID NO: 460, SEQ ID NO: 461, SEQ ID NO: 462,
 SEQ ID NO: 463, SEQ ID NO: 464, SEQ ID NO: 465, SEQ ID
 8860 NO: 466, SEQ ID NO: 467, SEQ ID NO: 468, SEQ ID NO: 469,
 SEQ ID NO: 470, SEQ ID NO: 471, SEQ ID NO: 472, SEQ ID
 NO: 473, SEQ ID NO: 474, SEQ ID NO: 475, SEQ ID NO: 476,

SEQ ID NO: 477, SEQ ID NO: 478, SEQ ID NO: 479, SEQ ID
 NO: 480, SEQ ID NO: 481, SEQ ID NO: 482, SEQ ID NO: 483,
 8865 SEQ ID NO: 645, SEQ ID NO: 646, SEQ ID NO: 647, SEQ ID
 NO: 648, SEQ ID NO: 649, SEQ ID NO: 650, SEQ ID NO: 651,
 SEQ ID NO: 652, SEQ ID NO: 653, SEQ ID NO: 654, SEQ ID
 NO: 655, SEQ ID NO: 656, SEQ ID NO: 657, SEQ ID NO: 658,
 SEQ ID NO: 659, SEQ ID NO: 660, SEQ ID NO: 661, SEQ ID
 8870 NO: 662, SEQ ID NO: 663, SEQ ID NO: 664, SEQ ID NO: 665,
 SEQ ID NO: 666, SEQ ID NO: 667, SEQ ID NO: 668, SEQ ID
 NO: 669, SEQ ID NO: 670, SEQ ID NO: 671, SEQ ID NO: 672,
 SEQ ID NO: 673, SEQ ID NO: 674, SEQ ID NO: 675, SEQ ID
 NO: 676, SEQ ID NO: 677, SEQ ID NO: 678, SEQ ID NO: 679,
 8875 SEQ ID NO: 680, SEQ ID NO: 681, SEQ ID NO: 682, SEQ ID
 NO: 683, SEQ ID NO: 684, SEQ ID NO: 685, SEQ ID NO: 686,
 SEQ ID NO: 687, SEQ ID NO: 688, SEQ ID NO: 877, SEQ ID
 NO: 878, SEQ ID NO: 879, SEQ ID NO: 880, SEQ ID NO: 881,
 SEQ ID NO: 882, SEQ ID NO: 883, SEQ ID NO: 884, SEQ ID
 8880 NO: 885, SEQ ID NO: 886, SEQ ID NO: 887, SEQ ID NO: 888,
 SEQ ID NO: 889, SEQ ID NO: 890, SEQ ID NO: 891, SEQ ID
 NO: 892, SEQ ID NO: 893, SEQ ID NO: 894, SEQ ID NO: 895,
 SEQ ID NO: 896, SEQ ID NO: 897, SEQ ID NO: 898, SEQ ID
 NO: 899, SEQ ID NO: 900, SEQ ID NO: 901, SEQ ID NO: 902,
 8885 SEQ ID NO: 903, SEQ ID NO: 904, SEQ ID NO: 905, SEQ ID
 NO: 906, SEQ ID NO: 907, SEQ ID NO: 908, SEQ ID NO: 909,
 SEQ ID NO: 910, SEQ ID NO: 911, SEQ ID NO: 912, SEQ ID
 NO: 913, SEQ ID NO: 914, SEQ ID NO: 915, SEQ ID NO: 916,
 SEQ ID NO: 917, SEQ ID NO: 918, SEQ ID NO: 919, SEQ ID
 8890 NO: 920, SEQ ID NO: 921, SEQ ID NO: 922, SEQ ID NO: 923,
 SEQ ID NO: 924, SEQ ID NO: 925, SEQ ID NO: 926, SEQ ID
 NO: 947, SEQ ID NO: 947, SEQ ID NO: 949, SEQ ID NO: 950,
 SEQ ID NO: 951, SEQ ID NO: 952, SEQ ID NO: 953, SEQ ID
 NO: 954, SEQ ID NO: 955, SEQ ID NO: 956, SEQ ID NO: 957,
 8895 SEQ ID NO: 958, SEQ ID NO: 959, SEQ ID NO: 960, SEQ ID
 NO: 961, SEQ ID NO: 962, SEQ ID NO: 963, SEQ ID NO: 964,

SEQ ID NO: 965, SEQ ID NO: 966, SEQ ID NO: 967, SEQ ID
 NO: 968, SEQ ID NO: 968, SEQ ID NO: 969, SEQ ID NO: 970,
 SEQ ID NO: 971, SEQ ID NO: 972, SEQ ID NO: 973, SEQ ID
 8900 NO: 1026, SEQ ID NO: 1027, SEQ ID NO: 1028, SEQ ID NO:
 1375, SEQ ID NO: 1376, SEQ ID NO: 1377, SEQ ID NO: 1378,
 SEQ ID NO: 1379, SEQ ID NO: 1410, SEQ ID NO: 1419, SEQ ID
 NO: 1420, SEQ ID NO: 1421, SEQ ID NO: 1422, SEQ ID
 NO: 1423, SEQ ID NO: 1424, SEQ ID NO: 1425, SEQ ID NO:
 8905 1488, SEQ ID NO: 1517, SEQ ID NO: 1516, SEQ ID NO: 1517,
 SEQ ID NO: 1538, SEQ ID NO: 1539, SEQ ID NO: 1550, SEQ ID
 NO: 1567, SEQ ID NO: 1568, SEQ ID NO: 1608, SEQ ID NO:
 1609, SEQ ID NO: 1610, SEQ ID NO: 1611, SEQ ID NO: 1628,
 SEQ ID NO: 1633, SEQ ID NO: 1634, SEQ ID NO: 1641, SEQ ID
 8910 NO: 1642, SEQ ID NO: 1644, SEQ ID NO: 1645, SEQ ID NO:
 1665, SEQ ID NO: 1676, and SEQ ID NO: 1681. These proteins
 or polypeptides are specific to O-157:H7, and significant
 homology to all data registered in gene data bank is found from
 determined information of amino acid sequence. Whereas,
 8915 their functions and the like are not known. However, as shown
 in table 1, a protein predicted to be a cell surface protein
 (membrane protein, especially, OMP, lipoprotein) in them or its
 gene (or nucleic-acid molecule) may be useful for production of
 an antibody, vaccine composition, diagnosis of O-157 infection
 8920 and the like. Furthermore, there is a possibility that they
 include a protein which has an important function in O-157, for
 example, transportation and metabolism of a substance,
 processing of nucleic acids, and relates to a regulatory element
 and pathogenicity. They are to be useful for diagnosis and
 8925 therapy of O-157 infection.

[0032]

3) Proteins comprising Insertion Sequence (IS)

These proteins or polypeptides are selected from a group
 comprising the following sequence list: SEQ ID NO: 133, SEQ
 8930 ID NO: 134, SEQ ID NO: 135, SEQ ID NO: 136, SEQ ID NO: 137,

SEQ ID NO: 138, SEQ ID NO: 139, SEQ ID NO: 140, SEQ ID
 NO: 141, SEQ ID NO: 142, SEQ ID NO: 143, SEQ ID NO: 144,
 SEQ ID NO: 145, SEQ ID NO: 146, SEQ ID NO: 147, SEQ ID
 NO: 148, SEQ ID NO: 149, SEQ ID NO: 150, SEQ ID NO: 151,
 8935 SEQ ID NO: 152, SEQ ID NO: 153, SEQ ID NO: 154, SEQ ID
 NO: 155, SEQ ID NO: 156, SEQ ID NO: 157, SEQ ID NO: 158,
 SEQ ID NO: 159, SEQ ID NO: 160, SEQ ID NO: 161, SEQ ID
 NO: 162, SEQ ID NO: 279, SEQ ID NO: 280, SEQ ID NO: 281,
 SEQ ID NO: 282, SEQ ID NO: 283, SEQ ID NO: 284, SEQ ID
 8940 NO: 285, SEQ ID NO: 286, SEQ ID NO: 287, SEQ ID NO: 288,
 SEQ ID NO: 289, SEQ ID NO: 290, SEQ ID NO: 291, SEQ ID
 NO: 292, SEQ ID NO: 293, SEQ ID NO: 294, SEQ ID NO: 295,
 SEQ ID NO: 296, SEQ ID NO: 297, SEQ ID NO: 298, SEQ ID
 NO: 299, SEQ ID NO: 300, SEQ ID NO: 301, SEQ ID NO: 302,
 8945 SEQ ID NO: 303, SEQ ID NO: 304, SEQ ID NO: 305, SEQ ID
 NO: 306, SEQ ID NO: 307, SEQ ID NO: 308, SEQ ID NO: 309,
 SEQ ID NO: 310, SEQ ID NO: 311, SEQ ID NO: 312, SEQ ID
 NO: 313, SEQ ID NO: 314, SEQ ID NO: 315, SEQ ID NO: 316,
 SEQ ID NO: 317, SEQ ID NO: 318, SEQ ID NO: 319, SEQ ID
 8950 NO: 320, SEQ ID NO: 321, SEQ ID NO: 322, SEQ ID NO: 323,
 SEQ ID NO: 324, SEQ ID NO: 325, SEQ ID NO: 326, SEQ ID
 NO: 327, SEQ ID NO: 328, SEQ ID NO: 329, SEQ ID NO: 330,
 SEQ ID NO: 331, SEQ ID NO: 332, SEQ ID NO: 333, SEQ ID
 NO: 334, SEQ ID NO: 335, SEQ ID NO: 336, SEQ ID NO: 1030,
 8955 SEQ ID NO: 1031, SEQ ID NO: 1032, SEQ ID NO: 1033, SEQ ID
 NO: 1034, SEQ ID NO: 1035, SEQ ID NO: 1036, SEQ ID NO:
 1037, SEQ ID NO: 1038, SEQ ID NO: 1039, SEQ ID NO: 1040,
 SEQ ID NO: 1041, SEQ ID NO: 1042, SEQ ID NO: 1043, SEQ ID
 NO: 1044, SEQ ID NO: 1045, SEQ ID NO: 1046, SEQ ID NO:
 8960 1047, SEQ ID NO: 1048, SEQ ID NO: 1049, SEQ ID NO: 1050,
 SEQ ID NO: 1051, SEQ ID NO: 1052, SEQ ID NO: 1053, SEQ ID
 NO: 1054, and SEQ ID NO: 1570. These proteins and their
 genes (or nucleic-acid molecules) are useful for detection and
 diagnosis of O-157 infection.

8965 [0033]

4) Proteins derived from phage:

These proteins or polypeptides are selected from a group comprising the following sequence list: SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 41, SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 47, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59, SEQ ID NO: 60, SEQ ID NO: 61, SEQ ID NO: 62, SEQ ID NO: 63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO: 69, SEQ ID NO: 70, SEQ ID NO: 71, SEQ ID NO: 72, SEQ ID NO: 73, SEQ ID NO: 74, SEQ ID NO: 75, SEQ ID NO: 76, SEQ ID NO: 77, SEQ ID NO: 78, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 81, SEQ ID NO: 82, SEQ ID NO: 83, SEQ ID NO: 84, SEQ ID NO: 85, SEQ ID NO: 86, SEQ ID NO: 87, SEQ ID NO: 88, SEQ ID NO: 89, SEQ ID NO: 90, SEQ ID NO: 91, SEQ ID NO: 92, SEQ ID NO: 93, SEQ ID NO: 94, SEQ ID NO: 95, SEQ ID NO: 96, SEQ ID NO: 97, SEQ ID NO: 98, SEQ ID NO: 99, SEQ ID NO: 100, SEQ ID NO: 101, SEQ ID NO: 102, SEQ ID NO: 103, SEQ ID NO: 104, SEQ ID NO: 105, SEQ ID NO: 106, SEQ ID NO: 107, SEQ ID NO: 108, SEQ ID NO: 109, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112, SEQ ID NO: 113, SEQ ID NO: 114, SEQ ID NO: 115, SEQ ID NO: 116, SEQ ID NO: 117, SEQ ID NO: 118, SEQ ID NO: 119, SEQ ID NO: 120, SEQ ID NO: 121, SEQ ID NO: 122, SEQ ID NO: 123, SEQ ID NO: 124, SEQ ID NO: 125, SEQ ID NO: 126, SEQ ID NO: 127, SEQ ID NO: 128, SEQ ID NO: 129, SEQ ID NO: 130, SEQ ID NO: 131, SEQ ID NO: 555, SEQ ID NO: 556, SEQ ID NO: 557, SEQ ID NO: 558, SEQ ID NO: 559, SEQ ID NO: 560, SEQ ID NO: 561, SEQ ID NO: 562, SEQ ID NO: 563, SEQ ID NO: 564, SEQ ID NO: 565, SEQ ID NO: 566, SEQ ID NO: 567, SEQ ID NO: 568, SEQ ID NO: 569, SEQ ID

NO: 570, SEQ ID NO: 571, SEQ ID NO: 572, SEQ ID NO: 573,
9000 SEQ ID NO: 574, SEQ ID NO: 575, SEQ ID NO: 576, SEQ ID
NO: 577, SEQ ID NO: 578, SEQ ID NO: 579, SEQ ID NO: 580,
SEQ ID NO: 581, SEQ ID NO: 582, SEQ ID NO: 583, SEQ ID
NO: 584, SEQ ID NO: 585, SEQ ID NO: 586, SEQ ID NO: 587,
SEQ ID NO: 588, SEQ ID NO: 589, SEQ ID NO: 590, SEQ ID
9005 NO: 591, SEQ ID NO: 592, SEQ ID NO: 593, SEQ ID NO: 594,
SEQ ID NO: 595, SEQ ID NO: 596, SEQ ID NO: 597, SEQ ID
NO: 598, SEQ ID NO: 599, SEQ ID NO: 600, SEQ ID NO: 601,
SEQ ID NO: 602, SEQ ID NO: 603, SEQ ID NO: 604, SEQ ID
NO: 605, SEQ ID NO: 606, SEQ ID NO: 607, SEQ ID NO: 608,
9010 SEQ ID NO: 609, SEQ ID NO: 610, SEQ ID NO: 611, SEQ ID
NO: 612, SEQ ID NO: 613, SEQ ID NO: 614, SEQ ID NO: 615,
SEQ ID NO: 616, SEQ ID NO: 617, SEQ ID NO: 618, SEQ ID
NO: 619, SEQ ID NO: 620, SEQ ID NO: 621, SEQ ID NO: 622,
SEQ ID NO: 623, SEQ ID NO: 624, SEQ ID NO: 625, SEQ ID
9015 NO: 626, SEQ ID NO: 627, SEQ ID NO: 628, SEQ ID NO: 629,
SEQ ID NO: 756, SEQ ID NO: 757, SEQ ID NO: 758, SEQ ID
NO: 759, SEQ ID NO: 760, SEQ ID NO: 761, SEQ ID NO: 762,
SEQ ID NO: 763, SEQ ID NO: 764, SEQ ID NO: 765, SEQ ID
NO: 766, SEQ ID NO: 767, SEQ ID NO: 768, SEQ ID NO: 769,
9020 SEQ ID NO: 770, SEQ ID NO: 771, SEQ ID NO: 772, SEQ ID
NO: 773, SEQ ID NO: 774, SEQ ID NO: 775, SEQ ID NO: 776,
SEQ ID NO: 777, SEQ ID NO: 778, SEQ ID NO: 779, SEQ ID
NO: 780, SEQ ID NO: 781, SEQ ID NO: 782, SEQ ID NO: 783,
SEQ ID NO: 784, SEQ ID NO: 785, SEQ ID NO: 786, SEQ ID
9025 NO: 787, SEQ ID NO: 788, SEQ ID NO: 789, SEQ ID NO: 790,
SEQ ID NO: 791, SEQ ID NO: 792, SEQ ID NO: 793, SEQ ID
NO: 794, SEQ ID NO: 795, SEQ ID NO: 796, SEQ ID NO: 797,
SEQ ID NO: 798, SEQ ID NO: 799, SEQ ID NO: 800, SEQ ID
NO: 801, SEQ ID NO: 802, SEQ ID NO: 803, SEQ ID NO: 804,
9030 SEQ ID NO: 805, SEQ ID NO: 806, SEQ ID NO: 807, SEQ ID
NO: 808, SEQ ID NO: 809, SEQ ID NO: 810, SEQ ID NO: 811,
SEQ ID NO: 812, SEQ ID NO: 813, SEQ ID NO: 814, SEQ ID

NO:815, SEQ ID NO: 1061, SEQ ID NO: 1062, SEQ ID NO: 1063,
 SEQ ID NO: 1064, SEQ ID NO: 1065, SEQ ID NO: 1066, SEQ ID
 9035 NO: 1067, SEQ ID NO: 1068, SEQ ID NO: 1069, SEQ ID NO:
 1070, SEQ ID NO: 1071, SEQ ID NO: 1072, SEQ ID NO: 1073,
 SEQ ID NO: 1074, SEQ ID NO: 1075, SEQ ID NO: 1076, SEQ ID
 NO:1077, SEQ ID NO: 1078, SEQ ID NO: 1079, SEQ ID NO:
 1080, SEQ ID NO: 1081, SEQ ID NO: 1082, SEQ ID NO: 1083,
 9040 SEQ ID NO: 1084, SEQ ID NO: 1085, SEQ ID NO: 1086, SEQ ID
 NO: 1087, SEQ ID NO: 1088, SEQ ID NO: 1089, SEQ ID NO:
 1090, SEQ ID NO: 1091, SEQ ID NO: 1092, SEQ ID NO: 1158,
 SEQ ID NO:1159, SEQ ID NO: 1160, SEQ ID NO: 1161, SEQ ID
 NO: 1162, SEQ ID NO: 1163, SEQ ID NO: 1164, SEQ ID NO:
 9045 1165, SEQ ID NO: 1166, SEQ ID NO: 1167, SEQ ID NO: 1168.
 SEQ ID NO: 1169, SEQ ID NO: 1170, SEQ ID NO: 1171, SEQ
 ID NO: 1172, SEQ ID NO: 1173, SEQ ID NO: 1174, SEQ ID
 NO: 1175, SEQ ID NO: 1176, SEQ ID NO: 1177, SEQ ID NO:
 1178, SEQ ID NO: 1179, SEQ ID NO: 1180, SEQ ID NO: 1181.
 9050 SEQ ID NO: 1182, SEQ ID NO: 1183, SEQ ID NO: 1184, SEQ
 ID NO: 1185, SEQ ID NO: 1186, SEQ ID NO: 1187, SEQ ID
 NO: 1188, SEQ ID NO: 1189, SEQ ID NO: 1190, SEQ ID NO:
 1259, SEQ ID NO: 1260, SEQ ID NO: 1261, SEQ ID NO: 1262,
 SEQ ID NO: 1263, SEQ ID NO: 1264, SEQ ID NO: 1265, SEQ ID
 9055 NO: 1266, SEQ ID NO: 1267, SEQ ID NO: 1268, SEQ ID NO:
 1269, SEQ ID NO: 1270, SEQ ID NO: 1271, SEQ ID NO: 1272,
 SEQ ID NO: 1273, SEQ ID NO: 1289, SEQ ID NO: 1290, SEQ ID
 NO: 1291, SEQ ID NO: 1292, SEQ ID NO: 1293, SEQ ID NO:
 1294, SEQ ID NO: 1295, SEQ ID NO: 1296, SEQ ID NO:1297,
 9060 SEQ ID NO: 1298, SEQ ID NO: 1299, SEQ ID NO: 1300, SEQ ID
 NO: 1301, SEQ ID NO: 1330, SEQ ID NO: 1331, SEQ ID NO:
 1332, SEQ ID NO: 1333, SEQ ID NO: 1334, SEQ ID NO: 1349,
 SEQ ID NO: 1350, SEQ ID NO: 1351, SEQ ID NO: 1352, SEQ ID
 NO: 1353, SEQ ID NO: 1354, SEQ ID NO: 1355, SEQ ID NO:
 9065 1356, SEQ ID NO: 1357, SEQ ID NO: 1358, SEQ ID NO: 1445,
 SEQ ID NO: 1446, SEQ ID NO: 1446, SEQ ID NO: 1447, 1448.

SEQ ID NO: 1449, SEQ ID NO:1490, SEQ ID NO: 1491, SEQ ID NO: 1492, SEQ ID NO: 1493, SEQ ID NO: 1509, SEQ ID NO: 1541, SEQ ID NO: 1542, SEQ ID NO: 1543, SEQ ID NO: 1544, 9070 SEQ ID NO: 1554, SEQ ID NO: 1572, SEQ ID NO: 1573, SEQ ID NO: 1574, SEQ ID NO: 1575, SEQ ID NO: 1581, SEQ ID NO: 1582, SEQ ID NO: 1583, SEQ ID NO: 1588, SEQ ID NO: 1589, SEQ ID NO: 1590, SEQ ID NO: 1597, SEQ ID NO: 1598, SEQ ID NO: 1623, SEQ ID NO: 1647, SEQ ID NO: 1648, SEQ ID NO: 9075 1650, SEQ ID NO: 1651, SEQ ID NO: 1653, 1654, SEQ ID NO: 1692, and SEQ ID NO:1693. These proteins and polypeptides are specific to O-157:H7 derived from phage. These proteins and their genes (or nucleic-acid molecule) are useful for detection and diagnosis of O-157 infection.

9080 [0034]

5) regulatory element:

These proteins or polypeptides are selected from the group comprising the following sequence list: SEQ ID NO: 1147, SEQ ID NO: 1148, SEQ ID NO: 1149, SEQ ID NO: 1150, SEQ ID NO: 1151, SEQ ID NO: 1152, SEQ ID NO: 1153, SEQ ID NO: 9085 1154, SEQ ID NO: 1155, SEQ ID NO: 1156, SEQ ID NO: 1192., SEQ ID NO: 1193, SEQ ID NO:1194, SEQ ID NO: 1335, SEQ ID NO: 1336, SEQ ID NO: 1337, SEQ ID NO: 1402, SEQ ID NO: 1403, SEQ ID NO: 1404, SEQ ID NO: 1405, SEQ ID NO: 1406, 9090 SEQ ID NO: 1407, SEQ ID NO: 1468, SEQ ID NO: 1512, SEQ ID NO: 1513, SEQ ID NO: 1514, SEQ ID NO: 1515, SEQ ID NO: 1585, SEQ ID NO: 1586, SEQ ID NO:1656, SEQ ID NO: 1657, SEQ ID NO: 1678, and SEQ ID NO: 1695. These proteins or polypeptides are O-157:H7 specific regulatory element and 9095 usable for development of a substance inhibiting expression of their genes. Such substance is useful for prevention and therapy of O-157 infection, and as a food additive. Furthermore, the protein and its gene (or nucleic-acid molecule) per se are useful for diagnosis and therapy of O-157 infection.

9100 [0035]

6) Proteins relating to fimbriae:

These proteins or polypeptides are selected from the group comprising the following sequence list: SEQ ID NO: 274, SEQ ID NO: 275, SEQ ID NO: 276, SEQ ID NO: 277, SEQ ID NO: 1195, SEQ ID NO: 1196, SEQ ID NO: 1197, SEQ ID NO: 1241, SEQ ID NO: 1242, SEQ ID NO: 1243, SEQ ID NO: 1244, SEQ ID NO: 1245, SEQ ID NO: 1246, SEQ ID NO: 1247, SEQ ID NO: 1248, SEQ ID NO: 1249, SEQ ID NO: 1250, SEQ ID NO: 1251, SEQ ID NO: 1252, SEQ ID NO: 1253, SEQ ID NO: 1254, SEQ ID NO: 1255, SEQ ID NO: 1256, SEQ ID NO: 1257, SEQ ID NO: 1427, SEQ ID NO: 1428, SEQ ID NO: 1429, SEQ ID NO: 1430, SEQ ID NO: 1431, SEQ ID NO: 1432, SEQ ID NO: 1433, SEQ ID NO: 1434, SEQ ID NO: 1435, SEQ ID NO: 1521, SEQ ID NO: 1522, SEQ ID NO: 1523, SEQ ID NO: 1524, SEQ ID NO: 1525, SEQ ID NO: 1548, SEQ ID NO: 1613, SEQ ID NO: 1614, SEQ ID NO: 1659, and SEQ ID NO: 1671. These proteins and their genes (or nucleic-acid molecules) are useful for production of antibody, vaccine composition, diagnosis of O-157 infection and the like. These proteins or polypeptides are available for development of a substance inhibiting expression of O-157:H7 specific gene. Such substance is useful for prevention and therapy of O-157 infection, and as a food additive. Furthermore, the protein and its gene (or nucleic-acid molecule) per se are useful for diagnosis and therapy of O-157 infection.

[0036]

7) Proteins relating to transportation of substance:

These proteins or polypeptides are selected from the group comprising the following sequence list: SEQ ID NO: 817, SEQ ID NO: 818, SEQ ID NO: 819, SEQ ID NO: 820, SEQ ID NO: 821, SEQ ID NO: 822, SEQ ID NO: 823, SEQ ID NO: 824, SEQ ID NO: 825, SEQ ID NO: 826, SEQ ID NO: 827, SEQ ID NO: 828, SEQ ID NO: 829, SEQ ID NO: 830, SEQ ID NO: 831, SEQ ID NO: 832, SEQ ID NO: 833, SEQ ID NO: 834, SEQ ID NO: 835, SEQ ID NO: 836, SEQ ID NO: 837, SEQ ID NO: 838,

9135 SEQ ID NO: 839, SEQ ID NO: 840, SEQ ID NO: 841, SEQ ID
NO: 842, SEQ ID NO: 843, SEQ ID NO: 844, SEQ ID NO: 1198,
SEQ ID NO: 1339, SEQ ID NO: 1340, SEQ ID NO: 1341, SEQ ID
NO: 1342, SEQ ID NO: 1343, SEQ ID NO: 1344, SEQ ID NO:
1345, SEQ ID NO: 1346, SEQ ID NO: 1347, SEQ ID NO: 1368,
9140 SEQ ID NO: 1369, SEQ ID NO: 1370, SEQ ID NO: 1371, SEQ ID
NO: 1458, SEQ ID NO: 1459, SEQ ID NO: 1461, SEQ ID NO: 14
62, SEQ ID NO: 1463, SEQ ID NO: 1464, SEQ ID NO: 1465, SEQ
ID NO: 1466, SEQ ID NO: 1507, and SEQ ID NO: 1679. These
proteins or polypeptides are regulatory elements specific to
9145 O-157:H7. These [proteins or polypeptides] are useful for
development of selection medium specific to O-157, or
development of a pharmaceutical agent selective to O-157, and
a strain comprising disruption in their genes may be useful as a
live attenuated vaccine. Furthermore, the protein and its gene
9150 (or nucleic-acid molecule) per se are useful for diagnosis and
therapy of O-157 infection.
[0037]

8) Proteins relating to synthesis of lipopolysaccharide:

These proteins or polypeptides are selected from the
9155 group comprising the following sequence list: EQ ID NO: 1533,
SEQ ID NO: 1534, SEQ ID NO: 1535, SEQ ID NO: 1536, SEQ ID
NO: 1395, SEQ ID NO: 1396, SEQ ID NO: 1397, SEQ ID NO:
1398, SEQ ID NO: 1399, SEQ ID NO: 1400, SEQ ID NO: 1412,
SEQ ID NO: 1413, SEQ ID NO: 1414, SEQ ID NO: 1415, SEQ ID
9160 NO: 1564, and SEQ ID NO: 1565. These proteins and their
gene (or nucleic-acid molecule) are especially useful for
production of antibody, vaccine composition, diagnosis of O-157
infection and the like. Furthermore, the protein and its gene
(or nucleic-acid molecule) per se are useful for diagnosis and
9165 therapy of O-157 infection.
[0038]

9) Proteins relating to metabolism:

These proteins or polypeptides are selected from the

group comprising the following sequence list: SEQ ID NO: 278,
 9170 SEQ ID NO: 690, SEQ ID NO: 691, SEQ ID NO: 692, SEQ ID
 NO: 693, SEQ ID NO: 694, SEQ ID NO: 695, SEQ ID NO: 696,
 SEQ ID NO: 697, SEQ ID NO: 698, SEQ ID NO: 699, SEQ ID
 NO: 700, SEQ ID NO: 701, SEQ ID NO: 702, SEQ ID NO: 703,
 SEQ ID NO: 704, SEQ ID NO: 705, SEQ ID NO: 706, SEQ ID
 9175 NO: 707, SEQ ID NO: 708, SEQ ID NO: 709, SEQ ID NO: 710,
 SEQ ID NO: 711, SEQ ID NO: 712, SEQ ID NO: 713, SEQ ID
 NO: 714, SEQ ID NO: 715, SEQ ID NO: 716, SEQ ID NO: 717,
 SEQ ID NO: 718, SEQ ID NO: 719, SEQ ID NO: 720, SEQ ID
 NO: 721, SEQ ID NO: 722, SEQ ID NO: 723, SEQ ID NO: 724,
 9180 SEQ ID NO: 725, SEQ ID NO: 726, SEQ ID NO: 727, SEQ ID
 NO: 728, SEQ ID NO: 729, SEQ ID NO: 730, SEQ ID NO: 731,
 SEQ ID NO: 1416, SEQ ID NO: 1417, SEQ ID NO: 1472, SEQ ID
 NO: 1552, SEQ ID NO: 1556, SEQ ID NO: 1557, SEQ ID NO:
 1616, SEQ ID NO: 1630, SEQ ID NO: 1631, SEQ ID NO: 1660,
 9185 SEQ IDNO: 1661, and SEQ ID NO: 1667. These proteins or
 polypeptides relate to O-157:H7 specific metabolism.
 Therefore, these [proteins or polypeptides] are useful for
 development of selection medium specific to O-157, or
 development of a pharmaceutical agent selective to O-157, and
 9190 a strain comprising disruption in their genes may be useful as a
 live attenuated vaccine. Moreover the protein or its gene (or
 nucleic-acid molecule) per se are useful for diagnosis and
 therapy of O-157 infection.
 [0039]

9195 10) Proteins relating DNA/RNA processing:

These proteins or polypeptides are selected from a group
 comprising the following sequence list: SEQ ID NO: 732, SEQ
 ID NO: 733, SEQ ID NO: 734, SEQ ID NO: 735, SEQ ID NO: 736,
 SEQ ID NO: 737, SEQ ID NO: 738, SEQ ID NO: 739, SEQ ID
 9200 NO: 740, SEQ ID NO: 741, SEQ ID NO: 742, SEQ ID NO: 743,
 SEQ ID NO: 744, SEQ ID NO: 745, SEQ ID NO: 1199, SEQ ID
 NO: 1200, SEQ ID NO: 1201, SEQ ID NO: 1202, SEQ ID NO:

1203, SEQ ID NO: 1204, SEQ ID NO: 1205, and SEQ ID
NO:1318. These [proteins or polypeptides] are useful for
9205 development of a pharmaceutical agent selective to O-157.
Furthermore, the protein and its gene (or nucleic-acid molecule)
per se are useful for diagnosis and therapy of O-157 infection.
[0040]

11) Proteins relating pathogenicity:

9210 These proteins or polypeptides are selected from a group
comprising the following sequence list: SEQ ID NO: 746, SEQ
ID NO: 747, SEQ ID NO: 748, SEQ ID NO: 749, SEQ ID NO: 750,
SEQ ID NO: 751, SEQ ID NO: 752, SEQ ID NO: 753, SEQ ID
NO: 754, SEQ ID NO: 845, SEQ ID NO: 846, SEQ ID NO: 847,
9215 SEQ ID NO: 848, SEQ ID NO: 849, SEQ ID NO: 850, SEQ ID
NO: 851, SEQ ID NO: 852, SEQ ID NO: 853, SEQ ID NO: 854,
SEQ ID NO: 855, SEQ ID NO: 856, SEQ ID NO: 857, SEQ ID
NO: 858, SEQ ID NO: 859, SEQ ID NO: 860, SEQ ID NO: 861,
SEQ ID NO: 862, SEQ ID NO: 863, SEQ ID NO: 864, SEQ ID
9220 NO: 865, SEQ ID NO: 866, SEQ ID NO: 867, SEQ ID NO: 868,
SEQ ID NO: 869, SEQ ID NO: 870, SEQ ID NO: 871, SEQ ID
NO: 872, SEQ ID NO: 873, SEQ ID NO: 874, SEQ ID NO: 875,
SEQ ID NO: 1129, SEQ ID NO: 1130, SEQ ID NO: 1131, SEQ ID
NO: 1132, SEQ ID NO: 1133, SEQ ID NO: 1134, SEQ ID NO:
9225 1135, SEQ ID NO: 1136, SEQ ID NO: 1137, SEQ ID NO: 1138,
SEQ ID NO: 1206, SEQ ID NO: 1207, SEQ ID NO: 1208, SEQ ID
NO: 1209, SEQ ID NO: 1210, SEQ ID NO: 1211, SEQ ID NO:
1310, SEQ ID NO: 1311, SEQ ID NO: 1312, SEQ ID NO: 1313,
SEQ ID NO: 1314, SEQ ID NO:1315, SEQ ID NO: 1316, SEQ ID
9230 NO: 1317, SEQ ID NO: 1321, SEQ ID NO: 1322, SEQ ID NO:
1323, SEQ ID NO: 1324, SEQ ID NO: 1325, SEQ ID NO: 1326,
SEQ ID NO: 1327, SEQ ID NO: 1328, SEQ ID NO: 1527, SEQ ID
NO: 1528, SEQ ID NO: 1529, SEQ ID NO: 1530, SEQ ID NO:
1531, SEQ ID NO: 1620, SEQ ID NO:1621, SEQ ID NO: 1674,
9235 and SEQ ID NO: 1686. These proteins or polypeptides are
relating to pathogenicity of O-157. Therefore, these [proteins

or polypeptides] are useful for development of a pharmaceutical agent selective to O-157 and the like. Furthermore, a strain comprising disruption in their genes may be useful as a live
9240 attenuated vaccine. Moreover, the protein or its gene (or nucleic-acid molecule) per se are useful for diagnosis and therapy of O-157 infection.

[0041]

12) Other proteins:

9245 These proteins or polypeptides are selected from a group comprising the following sequence list: SEQ ID NO: 1014, SEQ ID NO: 1015, SEQ ID NO: 1016, SEQ ID NO: 1017, SEQ ID NO: 1018, SEQ ID NO: 1019, SEQ ID NO: 1020, SEQ ID NO: 1021, SEQ ID NO: 1022, SEQ ID NO: 1023, SEQ ID NO: 1024, SEQ ID
9250 NO: 1025, SEQ ID NO: 1139, SEQ ID NO: 1140, SEQ ID NO: 1141, SEQ ID NO: 1142, SEQ ID NO: 1143, SEQ ID NO: 1144, SEQ ID NO: 1145, SEQ ID NO: 1146, SEQ ID NO: 1319, SEQ ID NO: 1320, SEQ ID NO: 1381, SEQ ID NO: 1382, SEQ ID NO: 1383, SEQ ID NO: 1384, SEQ ID NO: 1385, SEQ ID NO: 1469,
9255 SEQ ID NO: 1470, SEQ ID NO: 1546, SEQ ID NO: 1592, SEQ ID NO: 1593, SEQ ID NO: 1687, and SEQ ID NO: 1689. These proteins and their genes (or nucleic-acid molecules) are useful for detection and diagnosis of O-157 infection.

[0042]

9260 According to a standard technique in the art, the polypeptide of the present invention or a fragment thereof may be produced by inserting the nucleic-acid molecule of the present invention which encodes [the polypeptide or fragment] into a suitable expression vector, introducing the obtained
9265 recombinant vector to suitable host cells, culturing the host cells, and subsequently, extracting a desired polypeptide or a fragment thereof from the cultured host cells. Therefore, the present invention also relates to a method of producing O-157:H7 specific polypeptide comprising a recombinant
9270 expression vector containing the nucleic-acid molecule of the

present invention as an inserted substance, host cells transformed with the vector, and cultivation of the host cells.
[0043]

9275 In order to produce O-157 specific polypeptide of the present invention or a fragment thereof by using a technique for recombination, any expression system, for example, eukaryotic cells such as mammalian cells comprising human insect cells, fungal cells, yeast cells and the like, as well as, prokaryotic cells, for example, such as *E. coli* cells and the like
9280 may be used. The procaryotic cells are any known bacterial cells in the art. The cells include, for example, species of *E. coli*, salmonella, Norcardia, *Corynebacterium*, *Campylobacter*, *Streptomyces* (for example, Sambrook, Fritsch & Maniatis, Molecular Cloning; Laboratory Manual 2nd Ed., 1989).
9285 Examples of mammalian cells include COS7 cells or CHO cells. In case of [using] these cells, useful conventional promoters may be used for expression in mammalian cells. It is preferable that, for example, immediate early promoter of Human cytomegalo-virus (HCMV) is used. In addition, as a
9290 promoter for gene expression in mammalian cells which can be used in the present invention, virus promoters such as Retrovirus, polyomavirus, adenovirus, simian virus 40(SV40) and the like, or promoters derived from mammalian cells such as Human peptide chain elongation factor 1 α (HEF-1 α) and the
9295 like may be used. As a replication origin (ori), an ori derived from SV40, polyomavirus, adenovirus, Bovine papillomavirus may be used. In addition, the expression vector may include a gene of phosphotransferase APT(3') II or I (neo) and the like as a selection marker.
9300 [0044]

It is preferable that the recombinant expression vectors of the present invention includes DNA sequences encoding various antibiotic resistance genes or other marker genes as selection marker genes. Example of the marker genes include

9305 anti-spectinomycin gene, ampicillin resistance gene, streptomycin resistance gene (streptomycin phosphotransferase (SPT) gene), neomycin phosphotransferase (NPTII) gene of resistance to kanamycin or geneticin, hygromycin phosphotransferase (HTP) gene of hygromycin resistance,
9310 thymidine kinase (TK) gene, *E. coli* xanthine guanine phosphoribosyltransferase (Ecogpt) gene, dihydrofolate reductase (DHFR) gene, β -glucuronidase gene, luciferase gene, β -galactosidase gene, peroxidase gene and the like.
[0045]

9315 In order to detect O-157, Oligonucleotide primers for PCR can be constructed by using O-157 specific sequence in the nucleic-acid molecule or the gene of the present invention to perform rapid diagnosis of O-157. Basically, all of the O-157 specific sequences may be useful for a method for the rapid
9320 diagnosis by PCR. Therefore, the present invention relates to a method for detection or diagnosis of O-157 infection using the above mentioned oligonucleotide primer. Furthermore, the oligonucleotide may be used as a hybridization probe. The length of oligonucleotide of the present invention is at least 8
9325 nucleotides, preferably, 15 or more nucleotides, but may be determined, as necessary, by reference of a standard technique in genetic engineering.
[0046]

In addition to a nucleic-acid molecule having O-157
9330 specific nucleic acid sequence, the present invention also relates to a nucleic acid sequence comprising O-157 specific mutation which is also present in other *E. coli* (for example, strain of K-12) and a method of using it. Such nucleic acid sequences include, for example, a nucleic acid sequence
9335 comprising a mutation in genes relating to decrease of availability of sorbitol and lack of β -glucuronidase activity.
[0047]

O-157 specific nucleic-acid molecule of the present

invention, a gene included in it, peptide and nucleic-acid
9340 sequence encoded by the gene are useful for diagnosis and/or
therapy of O-157 infection and prevention of symptom occurred
by the infection. They can also be used for detection of the
presence of O-157 in a sample and classification of its strain.
Furthermore, they can also be used for screening of useful
9345 compounds for prevention and/or therapy of O-157 infection and
symptom occurred by the infection.
[0048]

The present invention also relates to an oligonucleotide
useful as a primer or a probe for detecting O-157 infection.
9350 Furthermore, the scope of the present invention includes a
vaccine composition including genes and/or polynucleotides of
the present invention, and a method for prevention and/or
therapy of O-157 infection and symptom occurred by the
infection.
9355 [0049]

Accordingly, the present invention relates to an
oligonucleotide or polynucleotide comprising a nucleotide
sequence constituted of at least 8 nucleotides in O-157 specific
nucleotide sequence set forth in the sequence lists, [a
9360 nucleotide sequence] comprising O-157 specific mutation, or a
complementary nucleic-acid sequence to the nucleic-acid
sequences. The present invention also relates to use of the
oligonucleotide or polynucleotide of the present invention used
as a hybridization probe or a PCR primer. The oligonucleotide
9365 used as a primer is comprised of at least 8 nucleotides,
preferably 15 nucleotides, more preferably at least 20 or more
nucleotides. The probe is comprised of at least 20 to 30
nucleotides. Nucleic acids used as a probe may be labeled by
using standard technique in the art.
9370 [0050]

Using the oligonucleotide or polynucleotide of the present
invention as a PCR primer, rapid diagnostic of O-157 may be

performed. Basically, all O-157 specific sequences may be useful for a method for rapid diagnosis by PCR. Therefore, the present invention relates to a method for detection or diagnosis of O-157 infection using the oligonucleotide primer.

[0051]

The present invention relates to a peptide vaccine formulation for prevention or therapy of O-157 infection comprising effective amount of, at least one kind of, O-157 specific polypeptides having amino acid sequence set forth in the sequence lists or fragments thereof. The vaccine formulation preferably includes a pharmaceutically acceptable carrier, for example, a known adjuvant in the art.

[0052]

The present invention also relates to a DNA vaccine formulation for prevention or therapy of O-157 infection comprising at least one of above mentioned O-157 specific polypeptides or polynucleotides encoding fragments thereof. The vaccine formulation preferably contains a pharmaceutically acceptable carrier, for example, an adjuvant and/or a transfection reagent and the like which are known in the art. The transfection reagent contains a liposome, a gold particle, and a cationic polymer suitable for transfecting a living cell with DNA vaccine. Use of the DNA vaccine against pathogenic bacteria is disclosed in, for example, an example of research of DNA vaccine, Han T. K. et al., DNA Cell Biol. 20(9), pp. 595-601, 2001; Miyaji E. N. et al., Vaccine 20(5-6), pp. 805-12, 2001, which is incorporated herein in its entirety by reference thereto.

[0053]

The present invention relates to a method of reducing the risk of O-157 infection in patients or a method for therapy [of the infection]. This method comprises administration of the vaccine formulation of the present invention to a patient so as to reduce the risk of O-157 infection or provide therapy of

infection.

[0054]

In other embodiment, the present invention relates to a
9410 method of producing the vaccine formulation of the present
invention. The method of producing the peptide vaccine
formulation includes combining at least one kind of O-157
specific polypeptide having the amino acid sequences set forth
in the sequence list and the fragments thereof with a
9415 pharmaceutically acceptable carrier.

[0055]

The method of producing the DNA vaccine formulation
includes inserting polynucleotide encoding at least one kind of
the polypeptides or the fragments thereof into the expression
9420 vector which can be expressed in a patient, and combining an
effective amount of the expression vector with a
pharmaceutically acceptable carrier. There is a possibility
that frequency of use of a codon is different between mammal
including human and *E. coli*. In this case, it is possible to
9425 improve the efficiency of translation of mRNA into a desired
polypeptide in a patient who should be treated or prevented
from O-157 infection by replacing codons of high frequency in
O-157 with codons of high frequency in mammal using a
standard technique in genetic engineering. A sequence such as
9430 intron A derived from cytomegalovirus may be included in the
expression vector to enhance the expression of desired
polypeptide. In the case where the DNA vaccine composition of
the present invention is administered to a human, the
recombinant expression vector is preferably [a vector] having a
9435 replication origin other than that of SV40. A sequence derived
from SV40 is not preferable, since there is a possibility that it
has carcinogenicity. The replication origins usable for this
purpose include, but not restricted to, replication origins
derived from, for example, other virus, prokaryotic cells,
9440 eukaryotic cells such as yeast cells or animal cells.

[0056]

The present invention also relates to an antibody selectively reacting with O-157 specific polypeptide or the fragment thereof. Anti-protein/anti-peptide, anti-serum or
9445 monoclonal antibody can be prepared according to a standard protocol (see, for example, Antibodies: A Laboratory Manual, Harlow & Lane ed., Cold Spring Harbor Press, 1988). In the present invention, the means of the term "antibody molecule" includes whole antibody, antibody fragments obtained by
9450 fragmentation using conventional technique, for example, Fab' and F(ab')₂ fragment, and single-chain Fv(scFv) obtained by a technique in genetic engineering. The antibody molecule of the present invention also includes an antibody fragment, a bispecific antibody comprising single-chain Fv or a chimera
9455 antibody. In this case, [the antibody molecule of the present invention] comprises two different antibodies against the same O-157 specific polypeptide, two antibodies recognizing different O-157 polypeptides, or one antibody against the polypeptide and one antibody recognizing an epitope which does not relate
9460 to O-157.

[0057]

A gene relating to O-157 specific metabolic function in O-157 specific genes is usable for development of novel medium for selection of O-157. Although, selection medium used at
9465 present is medium using comparatively specific property of O-157 such as decrease of availability of sorbitol, lack of β -glucuronidase activity, an ability of resistance to tellurite, there is a possibility that further specific [property] to O-157 is present in the genes of metabolic system found in the present
9470 invention. Such property is preferable for selection of O-157, preferably, is combined with decrease of availability of sorbitol, lack of β -glucuronidase activity and/or an ability of resistance to tellurite

[0058]

9475 A polypeptide relating to pathogenicity of O-157, a
bacterial surface protein, a regulatory protein, a protein
relating to metabolic system and a nucleic-acid molecule
encoding this [protein] is useful for development of a
pharmaceutical agent which selectively inhibits expression of
9480 pathogenicity of O-157. Therefore, the present invention
includes a method of searching or screening of a pharmaceutical
agent useful for prevention and/or therapy of symptom relating
to O-157. According to the method of the present invention,
novel preventive agent and/or therapeutic agent for symptom
9485 relating to O-157 may be provided.
[0059]

In addition, it may be performed to produce a
recombinant protein from a gene relating to pathogenicity
shown by the present invention, especially novel toxin, to
9490 analyse a function of the toxin, and to search inhibitor of the
toxin. Therefore, the present invention relates to a method of
searching or screening of inhibitor against the novel toxin.
Furthermore, it is possible to determine conformation on the
basis of a purified protein and information of amino acid
9495 sequence thereof and to design and synthesise the inhibitory
substances using computer. These inhibitory substances will
be not only an therapeutic agent of completely different type
from conventional antibiotics, but also be a food additive
selectively inhibiting growth of O-157.

9500 [0060]

In addition, the O-157 specific pathogenic gene, the gene
of bacterial cell surface protein and the regulatory gene of the
present invention may [be used for] developing a live
attenuated vaccine by preparing a disruptant thereof.
9505 Furthermore, a live attenuated vaccine may also be produced by
cloning dysfunctional gene corresponding to them into other
vaccine strain.

[0061]

On the other hand, a gene encoding an essential
9510 metabolic function for proliferation of O-157 in vivo or in vitro
or a regulatory gene may be [used for] preparing a mutant
which can proliferate under a specific condition in laboratory,
but cannot proliferate in mammalian living body including
human by preparing a strain comprising gene disruption in
9515 their genes. Such strain is useful as a live attenuated vaccine.
[0062]

In an embodiment of the present invention, a DNA
microarray or DNA chip includes a part or all of the nucleic
acid sequence or gene of the present invention. Preferably,
9520 there is provided a DNA chip or a method for producing the
DNA chip, wherein the DNA chip comprises

(a) a nucleotide sequence which is selected from a group
comprising the following SEQ IDs or a partial sequence thereof:
SEQ IDNO:1, SEQ ID NO:132, SEQ ID NO:244, SEQ ID NO:
9525 337, SEQ ID NO:410,SEQ ID NO:484, SEQ ID NO:554, SEQ ID
NO:630, SEQ ID NO:689, SEQ ID NO:755, SEQ ID NO:816,
SEQ ID NO:876, SEQ ID NO:927, SEQ ID NO:978,SEQ ID NO:
1013, SEQ ID NO:1029, SEQ ID NO:1055, SEQ ID NO:1060,
SEQID NO:1093, SEQ ID NO:1128, SEQ ID NO:1157, SEQ ID
9530 NO:1191, SEQ ID NO:1212, SEQ ID NO:1240, SEQ ID NO:1258,
SEQ ID NO:1274, SEQ ID NO:1288, SEQ ID NO:1302, SEQ ID
NO:1309, SEQ ID NO:1321, SEQ ID NO:1329,SEQ ID NO:1338,
SEQ ID NO:1348, SEQ ID NO:1359, SEQ ID NO:1366, SEQID
NO:1374, SEQ ID NO:1380, SEQ ID NO:1386, SEQ ID NO:1394,
9535 SEQ IDNO:1401, SEQ ID NO:1408, SEQ ID NO:1411, SEQ ID
NO:1418, SEQ ID NO:1426, SEQ ID NO:1436, SEQ ID NO:1443,
SEQ ID NO:1450, SEQ ID NO:1457, SEQ ID NO:1460, SEQ ID
NO:1467, SEQ ID NO:1471, SEQ ID NO:1473, SEQ ID NO:1478,
SEQ ID NO:1487, SEQ ID NO:1489, SEQ ID NO:1494, SEQ
9540 IDNO:1499, SEQ, ID NO:1501, SEQ ID NO:1506, SEQ ID NO:
1508, SEQ ID NO:1510, SEQ ID NO:1511, SEQ ID NO:1516,
SEQ ID NO:1520, SEQ ID NO:1526, SEQ ID NO:1532, SEQ ID

NO:1537, SEQ ID NO:1540, SEQ ID NO:1545,SEQ ID NO:1547,
 SEQ ID NO:1549, SEQ ID NO:1551, SEQ ID NO:1553, SEQID
 9545 NO:1555, SEQ ID NO:1558, SEQ ID NO:1563, SEQ ID NO:1566,
 SEQ ID NO:1569, SEQ ID NO:1571, SEQ ID NO:1576, SEQ ID
 NO:1580, SEQ ID NO:1584, SEQ ID NO:1587, SEQ ID NO:1591,
 SEQ ID NO:1594, SEQ ID NO:1596,SEQ ID NO:1599, SEQ ID
 NO:1601, SEQ ID NO:1603, SEQ ID NO:1604, SEQID NO:1605,
 9550 SEQ ID NO:1607, SEQ ID NO:1612, SEQ ID NO:1615, SEQ
 IDNO:1617, SEQ ID NO:1619, SEQ ID NO:1622, SEQ ID NO:
 1624, SEQ ID NO:1626, SEQ ID NO:1627, SEQ ID NO:1629,
 SEQ ID NO:1632, SEQ ID NO:1635, SEQ ID NO:1636, SEQ ID
 NO:1637, SEQ ID NO:1639, SEQ ID NO:1640, SEQ ID NO:1643,
 9555 SEQ ID NO:1646, SEQ ID NO:1649, SEQ ID NO:1652, SEQ
 IDNO:1655, SEQ ID NO:1658, SEQ ID NO:1660, SEQ ID NO:
 1662, SEQ ID NO:1664, SEQ ID NO:1666, SEQ ID NO:1668,
 SEQ ID NO:1669, SEQ ID NO:1670, SEQ ID NO:1672, SEQ ID
 NO:1673, SEQ ID NO:1675, SEQ ID NO:1677,SEQ ID NO:1680,
 9560 SEQ ID NO:1682, SEQ ID NO:1683, SEQ ID NO:1685, SEQID
 NO:1688, SEQ ID NO:1690, SEQ ID NO:1691, SEQ ID NO:1694,
 SEQ ID NO:1696, SEQ ID NO:1699, SEQ ID NO:1700, SEQ ID
 NO:1701, SEQ ID NO:1704, SEQ ID NO:1705, SEQ ID NO:1706,
 SEQ ID NO:1707, SEQ ID NO:1708,SEQ ID NO:1709, SEQ ID
 9565 NO:1710, SEQ ID NO:1711, SEQ ID NO:1712, SEQID NO:1713,
 SEQ ID NO:1715, SEQ ID NO:1716, SEQ ID NO:1717, SEQ
 IDNO:1718,, SEQ ID NO:1719, SEQ ID NO:1720, SEQ ID NO:
 1721, SEQ ID NO:1722, SEQ ID NO:1723, SEQ ID NO:1724,
 SEQ ID NO:1725, SEQ ID NO:1726, SEQ ID NO:1727, SEQ ID
 9570 NO:1728, SEQ ID NO:1729, SEQ ID NO:1730,SEQ ID NO:1731,
 SEQ ID NO:1732, SEQ ID NO:1733, SEQ ID NO:1734, SEQID
 NO:1735, SEQ ID NO:1736, SEQ ID NO:1737, SEQ ID NO:1738,
 SEQ ID NO:1739, SEQ ID NO:1740, SEQ ID NO:1741, SEQ ID
 NO:1742. SEQ ID NO:1743, SEQ ID NO:1744, SEQ ID NO:1745,
 9575 SEQ ID NO:1746, SEQ ID NO:1747,SEQ ID NO:1748, SEQ ID
 NO:1749, SEQ ID NO:1750, SEQ ID NO:1751, SEQID NO:1752,

SEQ ID NO:1753, SEQ ID NO:1754, SEQ ID NO:1755, SEQ
 IDNO:1756, SEQ ID NO:1757, SEQ ID NO:1758, SEQ ID NO:
 1759, SEQ ID NO:1760, SEQ ID NO:1761, SEQ ID NO:1762.
 9580 SEQ ID NO:1763, SEQ ID NO:1764, SEQ ID NO:1765, SEQ ID
 NO:1766, SEQ ID NO:1767, SEQ ID NO:1768, SEQ ID NO:1769,
 SEQ ID NO:1770, SEQ ID NO:1771, SEQ ID NO:1772, SEQ
 IDNO:1773, SEQ ID NO:1774, SEQ ID NO:1775, SEQ ID NO:
 1776, SEQ ID NO:1777, SEQ ID NO:1778, SEQ ID NO:1779,
 9585 SEQ ID NO:1780, SEQ ID NO:1781, SEQ ID NO:1782, SEQ ID
 NO:1783, SEQ ID NO:1784, SEQ ID NO:1785,SEQ ID NO:1786,
 SEQ ID NO:1787, SEQ ID NO:1788, SEQ ID NO:1789, SEQID
 NO:1790, SEQ ID NO:1791, SEQ ID NO:1792, SEQ ID NO:1793,
 SEQ ID NO:1794, SEQ ID NO:1795, SEQ ID NO:1796, SEQ ID
 9590 NO:1797, SEQ ID NO:1798, SEQ ID NO:1799, SEQ ID NO:1800,
 SEQ ID NO:1801, SEQ ID NO:1802,SEQ ID NO:1803, SEQ ID
 NO:1804, SEQ ID NO:1805, SEQ ID NO:1806, SEQID NO:1807,
 SEQ ID NO:1808, SEQ ID NO:1809, SEQ ID NO:1810, SEQ
 IDNO:1811, SEQ ID NO:1812, SEQ ID NO:1813, SEQ ID NO:
 9595 1814, SEQ ID NO:1815, SEQ ID NO:1816, SEQ ID NO:1817,
 SEQ ID NO:1818, SEQ ID NO:1819, SEQ ID NO:1820, SEQ ID
 NO:1821, SEQ ID NO:1822, SEQ ID NO:1823, SEQ ID NO:1824,
 SEQ ID NO:1825, SEQ ID NO:1826, SEQ ID NO:1827, SEQ
 IDNO:1828, SEQ ID NO:1829, SEQ ID NO:1830, SEQ ID NO:
 9600 1831, SEQ ID NO:1832, SEQ ID NO:1833, SEQ ID NO:1834,
 SEQ ID NO:1835, SEQ ID NO:1836, SEQ ID NO:1837, SEQ ID
 NO:1838, SEQ ID NO:1839, SEQ ID NO:1840,SEQ ID NO:1841,
 SEQ ID NO:1842, SEQ ID NO:1843, SEQ ID NO:1844, SEQID
 NO:1845, SEQ ID NO:1846, SEQ ID NO:1847, SEQ ID NO:1848,
 9605 SEQ ID NO:1849, SEQ ID NO:1850, SEQ ID NO:1851, SEQ ID
 NO:1852, SEQ ID NO:1853, SEQ ID NO:1854, SEQ ID NO:1855,
 SEQ ID NO:1856, SEQ ID NO:1857,SEQ ID NO:1858, SEQ ID
 NO:1859, SEQ ID NO:1860, SEQ ID NO:1861, SEQID NO:1862,
 SEQ ID NO:1863, SEQ ID NO:1864, SEQ ID NO:1865, および
 9610 SEQ ID NO:1866,

, and/or (b) an oligonucleotide or polynucleotide comprising complementary sequence to the sequences set forth in (a). Such DNA microarray or DNA chip may be produced using the nucleic acid sequence or gene of the present invention
9615 by a standard technique in the art (see, for example, "DNA Microarrays: A Practical Approach", Mark Schena, ed. Oxford: Oxford University Press, 1999, ISBN 0-19-963777-8; "Microarray Biochip Technology", Mark Schena, ed. Natick, MA: Eaton Publishing, 2000, ISBN 1-881299-37-6; "DNA Arrays: Methods and Protocols", Jang B. Rampal, ed. Totowa, NJ: Humana Press, 2001, ISBN 0-89603-822-X). The DNA microarray or DNA chip is usable for analysis of a function of O-157 specific gene, classification of strain of O-157, search of the presence or absence of a gene which is similar to that of
9620 other strain of O-157 or other type of strain of large intestine. The classification of strain using DNA array is disclosed in, for example, Salama N. et al., Proc. Natl. Acad. Sci. U A. 97(26), pp. 14668-73, 2000. A technique for detecting a pathogenic bacterium by using the DNA array is disclosed in, for example,
9625 Call D. R. et al., IntJ Food Microbiol, 67(1-2), pp.71-80, 2001. A technique for analysing expression of a gene using DNA array is disclosed in, for example, Harrington C. A. et al., Curr. Opin. Microbiol. 3(3), pp.285-91, 2000. The entity of these documents is incorporated herein by reference.
9630

9635 [0063]

Definition

In the present invention, the terms "O-157 specific" and "specific to O-157:H7" means that [a substance is] absent from nonpathogenic E. coli K-12, but is present in O-157 (or
9640 O-157:H7). Therefore, there is a possibility that, sometimes, the same substance or the similar substance is present in other type of E. coli or other strain of bacteria.

[0064]

In the present invention, the term "hybridize" means that

9645 hybridization is performed under a stringent condition, for
example, in 0.5xSSC solution, at 65°C or equivalent condition.
[0065]

The term "(cell) surface protein" used herein means all
proteins capable of approaching to the surface, such as inner
9650 membrane and outer membrane proteins, proteins which bind to
cell wall, and secretory proteins.
[0066]

The term "open reading frame (ORF)" means a region in
nucleic acids encoding a polypeptide or a part thereof. The
9655 ORF can be determined by [a region] from initiation codon to
termination codon or from termination codon to termination
codon.
[0067]

The term "coding sequence" used herein means nucleic
9660 acids which is transcribed into mRNAs and/or translated into
polypeptides in case where the coding sequence is placed under
regulation of a suitable regulatory sequence. The coding
sequence includes, but not restricted to, mRNA, synthetic DNA,
and recombinant nucleic acid sequence.
9665 [0068]

In the present application, the terms "a part" or
"fragment" of polypeptide means an oligopeptide or polypeptide
comprising at least 10 amino acid residues, preferably at least
20 amino acid residues, more preferably at least 40 amino acid
9670 residues. Furthermore, the terms "a part" or "fragment" of
nucleotide sequence also mean a nucleotide sequence
comprising at least 20 or more nucleotides, preferably 50 or
more nucleotides.
[0069]

9675 In the present application, the term "expression
regulatory element" or "expression regulatory sequence" means
a sequence capable of inducing and/or regulating expression of
a coding sequence or ORF linked thereto. The term "linked in

their action" means that above mentioned expression regulatory
9680 element or [expression regulatory] sequence is linked to a
coding sequence or ORF in the manner where the coding
sequence or ORF can be transcribed.
[0070]

In the present invention, metabolism of a substance
9685 means any aspects including, expression, function, action or
regulation of a substance. The metabolism of a substance
includes modification of a substance, for example, modifying
the substance with a covalent bond or a noncovalent bond. The
metabolism of a substance includes modification in other
9690 substances induced by the substance, for example, modifying
the other substances with a covalent bond or a noncovalent
bond. The metabolism of substance also includes alteration in
distribution of the substance. The metabolism of a substance
includes alteration in distribution of other substance induced
9695 by the substance.
[0071]

In the present invention, transportation of a substance
means transportation of a substance from extracellular space to
intracellular space, transportation of a substance within a cell,
9700 and secretion and release of a substance to extracellular space.
[0072]

On carrying out the present invention, common
techniques in the art may be applied unless particularly
otherwise indication. Such techniques are disclosed in
9705 Sambrook, Fritsch & Maniatis, Molecular Cloning; Laboratory
Manual 2nd Ed. (1989); DNA Cloning, Volume (D.N. Glover Ed.
1985); Oligonucleotide Synthesis (M.J. Gait Ed. 1984); Nucleic
Acid Hybridization (B.D. Hames & S.J. Higgins Ed. 1984);
Methods in Enzymology (Academic Press, Inc.), Vol. 154 & Vol.
9710 155 (Wu & Grossman ed.) and PCR-A Practical Approach
(McPherson, Quirke & Taylor, ed. 1991).
[0073]

The nucleic acid molecule of the present invention may be directly obtained from the DNA of above mentioned O157:H7
9715 Sakai by using Polymerase Chain Reaction (PCR). Reliability of amplified product may be checked by a conventional method for determining sequence. A clone having a desired sequence set forth in the present invention may also be obtained by library screening using PCR or, by library screening using a
9720 synthetic oligonucleotide probe to library colonies or plaques lifted onto a filter, as known in the art (for example, Sambrook et al., Molecular Cloning, A Laboratory Manual 2nd edition, 1989, Cold Spring Harbor Press, NY). Nucleic acids encoding the polypeptides specific to O-157:H7 can also be obtained.
9725 [0074]

The nucleic acids of the present invention may also be chemically synthesized by using standard technique. Various methods for chemical synthesis of poly-deoxynucleotide are known (see, for example, Itakura et al., U.S. Patent
9730 No.4,598,049; Caruthers et al., U.S. Patent No.4,458,066; and Itakura et al., U.S. Patent No.4,401,796 and No.4,373,071, incorporated by reference herein).
[0075]

The present invention is explained by, but not restricted to, the following examples.
9735 [0076]

[Examples]

Example 1 : Determination of genomic nucleotide sequence of enterohemorrhagic pathogenic E. coli O-157:H7

9740 Whole nucleotide sequences on the chromosome of enterohemorrhagic E.coli O157:H7 were determined to identify regions and nucleic acid sequences which were specific to O157:H7, but absent from nonpathogenic E. coli K-12. The following strain was used in the Example: O157:H7 (RIMD
9745 0509952) which was isolated from a patient suffered from typical hemorrhagic enteritis during outbreak of O157:H7

infection which was occurred in mainly Sakai, Osaka, 1996. The strain has been stored in Research Center for Emerging Infectious Diseases, Research Institute for Microbial Diseases, Osaka University, and procedure for registration to ATCC (American Type Culture Collection) is now proceeding. The strain was cultured to prepare genomic DNA according to a conventional method. Random shotgun library comprising insertion of DNA fragment of 1-2 kbp in size was prepared to determining sequences of 50105 clones. With respect to 19969 clones among them, sequences at both end of the inserted fragment were determined (whole genome random shotgun sequencing). In addition, a library of lambda phage comprising inserted DNA fragments of about 20kbp was prepared to determine whole sequences of each of 86 clones individually. Assembly of the data of whole sequence which was obtained by using Phred/Phrap/consed was performed to obtain 111 contigs of 1 kbp or more. Finally, gap region between each of the contigs was amplified by using PCR and sequences of each PCR products were determined to determine the whole nucleotide sequences on chromosome of O157:H7. Then, the nucleotide sequence was analyzed by using a program such as Genome Gambler version 1.41, GLIMMER 2.01, BLAST and etc. to determine protein coding region. Furthermore, chromosomal sequence of O157:H7 was compared to chromosomal sequence of nonpathogenic *E.coli* K-12 (MG1655) using MUMmer Program to identify all regions of 20bp or more which is absent from K-12, but specifically present in O-157:H7. Determined chromosomal nucleotide sequences of O157:H7 has been registered in gene data bank DDBJ on 26 June, 2000 as Accession number: BA000007.

[0077]

Example 2: Detection of O-157 by PCR

On the basis of a nucleotide sequence of the Urease gene specifically present in O-157 Sakai, oligonucleotide primers

capable of amplifying Urease gene were synthesized. Detection of O-157 specific Urease gene by PCR was performed according to a conventional method using O-157 Sakai or various strains of *E. coli* as samples and the synthesized primers. As a result, 9785 the Urease gene was merely detected in enterohemorrhagic *E. coli* including O-157, whereas, not in other types of *E. coli*. In addition, it was found that the Urease gene was present in O-157 and closely related strains thereof, and it was shown that the primers were usable for rapid identification and 9790 diagnosis of O-157.

[0078]

Example 3: Molecular epidemiology of O-157 by PCR

On the basis of the nucleotide sequence information of O-157 Sakai, oligonucleotide primers specific to O-157 were 9795 synthesized. Examining a number of other strains of O-157 by PCR using the primers, it was found that a specific band was detected in some strains, whereas not in others. This result indicates the presence or absence of a specific sequence depending on the strains and makes it possible to identify 9800 regions containing a lot of differences between the strains. It was made possible to classify the strains of O-157 by using the primers amplifying the regions.

[0079]

Example 4: Applying the nucleotide sequence to Diagnosis

9805 The genetic information obtained in the Example 1 was analysed, resulting in suggestion of the presence of salicylic acid degradation gene specifically present in O-157. Accordingly, medium comprising salicylic acid as a carbon source was prepared by using a function of the salicylic acid 9810 degradation gene to perform a culture experiment. As a result, it is shown that O-157 could proliferate in the medium and there was a possibility that O-157 could be selectively isolated using the medium.

[0080]

9815 Example 5: Applying a nucleotide sequence to diagnosis

The genetic information obtained in the Example 1 was analyzed, resulting in finding the presence of mutations in coding sequence of β -glucuronidase gene (SEQ ID NO:1865) and coding sequence of gene of specific PTS enzyme IIB and IIC (SEQ ID NO:1866). The mutations included frame-shift mutation. Accordingly, an oligonucleotide primer against these mutations was synthesized to detect O-157 and other strain by PCR using the primer. As a result, absence of β -glucuronidase and decrease of availability of sorbitol could be confirmed without cultivation of the bacteria. A primer for detecting tellurite resistance gene was synthesized to perform PCR in the same way. As a result, a mutation in the tellurite resistance gene could be detected. Furthermore, by PCR using a combination of the three types of primers, higher accuracy results of diagnosis was obtained. According to Example 5, it was shown that these primers may be applied to rapid diagnosis of O-157

9825 [0081]

Example 6: Expression of a polypeptide

9835 A gene of a bacterial surface protein which was specifically present in O-157 was cloned to construct a system for mass production of a recombinant protein. The recombinant protein was purified using this system to construct a system for determining an antibody in patient's serum. It was shown that, this system was usable for serodiagnosis of O-157.

9840 [0082]

Example 7: Application of a nucleotide sequence for diagnosis

Based on the information of nucleotide sequence determined in Example 1, a toxin gene found newly was cloned to construct a system for mass production of a recombinant protein. The recombinant protein may be purified using this system, analyzed for a function of the toxin and searched for an

9845

inhibitor thereof. Based on the information of the purified
9850 protein and an amino acid sequence thereof, it is possible to
determine [their] conformation to design an inhibitory
substance and to synthesize [the inhibitory substance]. The
inhibitory substance will be a therapy agent of different type
from conventional antibiotics

9855 [0083]

Example 8: DNA Vaccine

A gene of a bacterial surface protein which was
specifically present in O-157 was cloned into a vaccine strain of
salmonella to confirm that the O-157 specific bacterial surface
9860 protein is expressed at surface of the vaccine strain of
salmonella. The vaccine strain is usable as a vaccine against
O-157.

[0084]

Example 9: Live attenuated vaccine

9865 A nucleic-acid molecule encoding a bacterial surface
protein which was specifically present in O-157 was inserted to
an expression vector suitable for salmonella to clone [the
expression vector] into a vaccine strain of attenuated
salmonella. Then, it was confirmed that the O-157 specific
9870 surface protein was expressed at surface of the vaccine strain
of salmonella. The vaccine strain is usable as a live vaccine
against O-157.

[0085]

Example 10: DNA microarray

9875 O-157 specific gene was amplified by PCR to prepare a
DNA chip according to a conventional method. mRNAs were
prepared from bacterial cells of O-157 which was cultured
under various culture conditions to analyse using the DNA chip.
As a result, it will be possible to perform various studies, such
9880 as [a study] of regulatory mechanism of expression of O-157
gene and [a study] of [confirming] whether a gene is expressed,
or not, under a certain condition.

[0086]

[Industrial applicability]

9885 The present invention provides a nucleotide sequence and
a polypeptide encoded thereby which are specific to
enterohemorrhagic E.coli O157:H7. These may be useful for
detection and/or therapy of infection. In addition, the present
invention provides a vaccine composition for preventing or
9890 treating O-157 infection. Furthermore, the present invention
has a possibility of providing a method of screening a novel
pharmaceutical agent and a food additive, and a method of
preventing and/or treating a pathosis relating to O-157.

9895

ABSTRACT

[Problems to be solved]

9900

Providing a nucleic-acid molecule, a polypeptide, genetic information thereof and a method of using them which may be useful for detection and therapy of enterohemorrhagic pathogenic-E. coli O-157:H7 infection.

[Means to solve the problem]

9905

Revealing genetic information of novel nucleic acid molecules specific to O-157, novel genes included the nucleic acid molecules, and novel polypeptides encoded by the genes.